STATISTICAL ANALYSIS OF ALFALFA GALAXIES: INSIGHTS IN GALAXY FORMATION & NEAR-FIELD COSMOLOGY

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Emmanouil Papastergis

August 2013

STATISTICAL ANALYSIS OF ALFALFA GALAXIES: INSIGHTS IN GALAXY FORMATION & NEAR-FIELD COSMOLOGY

Emmanouil Papastergis, Ph.D.

Cornell University 2013

The Arecibo Legacy Fast ALFA (ALFALFA) survey is a blind, extragalactic survey in the 21cm emission line of atomic hydrogen (HI). Presently, sources have been cataloged over $\approx 4\,000$ deg^2 of sky ($\sim$60% of its final area), resulting in the largest HI-selected sample to date. We use the rich ALFALFA dataset to measure the statistical properties of HI-bearing galaxies, such as their mass distribution and clustering characteristics. These statistical distributions are determined by the properties of dark matter on galactic scales, and by the complex baryonic processes through which galaxies form over cosmic time. As a result, detailed studies of these distributions can lead to important insights in galaxy formation & evolution and near-field cosmology.

In particular, we measure the space density of HI-bearing galaxies as a function of the width of their HI profile (i.e. the velocity width function of galaxies), and find substantial disagreement with the distribution expected in a lambda cold dark matter (ΛCDM) universe. In particular, the number of galaxies with maximum rotational velocities $v_{rot} \approx 35\ \mathrm{km\,s^{-1}}$ (as judged by their HI velocity width) is about an order of magnitude lower than what predicted based on populating ΛCDM halos with modeled galaxies. We identify two possible solutions to the discrepancy: First, an alternative dark matter scenario in which the formation of low-mass halos is heavily suppressed (e.g. a warm dark matter universe with keV-scale dark matter particles). Secondly, we consider the

possibility that rotational velocitites of dwarf galaxies derived from HI velocity widths may systematically underestimate the true mass of the host halo, due to the shape of their rotation curves.In this latter scenario, quantitative predictions for the internal kinematics of dwarf galaxies can be made, which can be checked in the future to probe the nature of dark matter.

Furthermore, we take advantage of the overlap of ALFALFA with the Sloan Digital Sky Survey (SDSS), to measure the number density of galaxies as a function of their "baryonic" mass (stars + atomic gas). In the context of a ΛCDM cosmological model, the measured distribution reveals that low-mass halos are heavily "baryon depleted", i.e. their baryonic-to-dark mass ratio is much lower than the cosmological value. These baryon deficits are usually attributed to stellar feedback (e.g. supernova-driven gas outflows), but the efficiency implied by our measurement is extremely high. Whether such efficient feedback can be accommodated in a consistent picture of galaxy formation is an open question, and remains one of the principle scientific drivers for hydrodynamic simulations of galaxy formation.

Lastly, we measure the clustering properties of HI-selected samples, through the two-point correlation function of ALFALFA galaxies. We find no compelling evidence for a dependence of clustering on HI mass, suggesting that the relationship between galactic gas mass and host halo mass is not tight. We furthermore find that HI galaxies cluster more weakly than optically selected ones, when no color selection is applied. However, SDSS galaxies with blue colors have very similar clustering characteristics with ALFALFA galaxies, both in real as well as in redshift space. On the other hand, HI galaxies cluster much more weakly than optical galaxies with red colors, and in fact "avoid" being located within $\approx$3 Mpc from the latter. By considering the clustering properties

of ΛCDM halos, we confirm our previous intuition for an M_{HI}-M_h relation with large scatter, and find that spin parameter may be a key halo property related to the gas content of present-day galaxies.

BIOGRAPHICAL SKETCH

Emmanouil (Manolis) Papastergis was born in the town of Merano, in the Alto Adige/Südtirol province of Italy. He mostly grew up in the small village of Soroni, on the island of Rhodos, Greece with his father Lefteris, his mother Pia and younger brother Stergos. During his school years he enjoyed looking at the beautiful night sky of rural Rhodos, and developed a passion for astronomy. After graduating from Soroni high school, he attended the Aristotle University of Thessaloniki, in Thessaloniki, Greece, where he studied Electrical & Computer Engineering with a specialization in Telecommunications. After graduating from college in 2007, he decided to follow his earlier passion and apply for graduate studies in astronomy in the United Sates. He was accepted in the Astronomy Department of Cornell University, where he joined the ALFALFA survey team, supervised by Professors Riccardo Giovanelli & Martha P. Haynes. Manolis spent most of his time as a graduate student in the "Camuy Cave", where he mostly worked on the statistical analysis of the galaxies detected by ALFALFA (his research was in essence a sophisticated version of "counting galaxies", as his family and friends were told). Notwithstanding the considerable amount of time spent on work breaks (during which he would distract his fellow grad students from their own research), he managed to graduate in August of 2013. After an exhausting application effort, he was offered a NOVA fellowship as a postdoctoral researcher at the Kapteyn Institute, University of Groningen, Netherlands.

This dissertation is dedicated to my parents, Lefteris and Pia, who made it possible for me to study astronomy. And to my grandfather, Gastone Mazzucato, who bought me my first astronomy book.

ACKNOWLEDGEMENTS

- Riccardo & Martha – When someone starts grad school it is easy to think that all advisors are as good as you. It took me some time to realize how lucky I was to be working for you (actually *with* you). Thank you for giving me the opportunity to join the ALFALFA team, and for letting me work on the aspects of the survey that I enjoyed the most. Also, thank you for keeping everything running so smoothly (I bet that making IDL figures in the future will take many more steps than just typing @egggeninit). Also, thank you for your "titanic effort" (as we say in Greece) assisting me during my postdoc application season. 38 out of 38 letters of recommendation submitted on time was no easy task!
- All members of the ALFALFA team – Thank you for observing, flagging, baselining, flatfielding, and extracting thousands of sources by hand. None of the work presented in this dissertation would have been possible if I did not have a fully processed catalog in my hands.
- Ann Martin – Thank you for teaching me how to do statistics with ALFALFA galaxies. All the work presented in this dissertation is based – directly or indirectly– on the methods you helped me learn. Thank you also for writing the most readable and detailed thesis I have ever encountered in my life; I would be very happy if this thesis came even somewhat close.
- Betsey, Shan, Greg, Mike, Luke – It was great sharing an office with you. I am sure you too enjoyed listening to my skype conversations with my mother and to my swearing at my computer in greek.
- Laura, Dan, Eva, Drew, Joyce – Thank you for putting up with me. I am sure that every time you heard me calling your name without any appar-

ent reason, you had visions of yourself holding a knife. Also, thank you Dan for answering my questions and for listening to the answers I gave to your questions; hopefully some of them were right.

- Andrea Cattaneo and Aldo Rodriguez-Puebla – I am very glad I had the chance to collaborate with you. A special thanks goes to Andrea for writing me endless letters of recommendation and for the wonderful hospitality at Aix.
- The National Science Foundation, Cornell University, NASA, the Onassis Foundation and the Brinson Foundation – Thank you for keeping me funded during all six years of my graduate career. We do not like to admit it, but astronomers rely on money as much as any other professionals.
- The owl blanket in the grad lounge – Thank you for keeping me company during my (almost daily) naps. If I am any successful in the future it will be mostly because of the time I spend asleep.
- Mary – I don't think I would have been able to go through grad school without your help. Thank you for keeping me sane during the stressful periods, and for encouraging me to always do my best at work (as opposed to giving up and spending my days laying on your couch). Very soon you'll be dating a doctor –unfortunately not the rich kind.
- Lefteris & Pia – Thank you for helping me in all imaginable ways during these past six years (they would have been five if I didn't spend so much time on skype with you). I am especially happy you managed to visit me for my graduation day and my B-exam. Very soon I'll be in the Netherlands and I promise to visit more often. And you'll have a son who is a doctor – unfortunately not the useful kind.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

The advent of large-scale galaxy surveys in recent years has revolutionized our understanding of how galaxies form and evolve, and has proven a powerful tool for determining the properties of our Universe. Most of the effort in this area has focused on optical surveys, where projects such as the Two-Degree Field Galaxy Redshift Survey (2dFGRS; Colless et al., 2001) and Sloan Digital Sky Survey (SDSS; York et al., 2000), have scanned large areas of the sky in a systematic fashion, measuring the light and spectra of $>$1 million galaxies. With such sample sizes, statistics have become an indispensable tool for learning about the present and past properties of galaxies.

For example, by measuring the distribution of galaxy stellar masses (i.e. the number density of galaxies as a function of their stellar mass, referred to as the stellar mass function of galaxies) one can learn that no galaxy is able to turn its "cosmic share" of baryons into stars. In fact, Milky Way-sized galaxies ($M_* \approx 10^{10.5}\ M_\odot$) are the most efficient in turning baryons into stars, albeit with an unimpressive star formation efficiency[1] of $\lesssim$30%. The situation is even bleaker for dwarf galaxies, having star formation efficiencies $\eta_* \lesssim 1\%$ at $M_* \approx 10^{7.5}\ M_\odot$ [e.g. Guo et al., 2010, Moster et al., 2010, Behroozi et al., 2010, Leauthaud et al., 2010].

The study of the clustering properties of galaxies, mostly through the measurement of the galaxy correlation function, constitutes yet another example where statistical analyses play a key role in furthering our knowledge of extra-

[1]The galaxy star formation efficiency, η_*, is defined as the ratio of the stellar mass of the galaxy to the total mass of its host halo, normalized to the cosmic baryon fraction $f_b \equiv \Omega_b/\Omega_m$. In other words, $\eta_* = (M_*/M_h)\ /\ f_b$, where $f_b \approx 0.16$.

galactic astronomy and cosmology. Galaxies have been found to cluster more strongly with increasing luminosity or stellar mass [e.g. Norberg et al., 2002, Zehavi et al., 2011, Li et al., 2012b, Marulli et al., 2013], giving us observational confirmation for the theoretical expectation that more massive galaxies are hosted by more massive halos. In addition, galaxies with red optical colors and early-type morphologies are found to cluster much more strongly than galaxies with blue colors and late-type morphologies [e.g. Norberg et al., 2002, Zehavi et al., 2011, Christodoulou et al., 2012], showing that red galaxies have a higher chance of being hosted by a subhalo, i.e. a bound dark matter structure that is located within the virial radius[2] of a larger halo.

In both preceding examples the cosmological model was assumed to be known, and the statistical properties of galaxies were used to learn about their formation and evolution in the cosmological context. However, similar techniques can be used in the reverse direction: galaxies can be used as tracers of the underlying distribution of matter, in order to test the concordance ΛCDM cosmological model and measure its parameters to high precision. For example, the large-scale correlation function of galaxies bears the imprint of pressure waves traveling through the cosmic plasma at around the time of recombination ($\approx 400\,000$ years after the big bang), the same ones that produce the characteristic anisotropy pattern in the Cosmic Microwave Background (CMB). This phenomenon manifests itself in a broad peak in the correlation function, referred to as the Baryon Acoustic Oscillation (BAO) feature, and is located at a well known physical length scale of ≈150 Mpc. The BAO peak can therefore be used

[2]The virial radius of a halo is a measure of its spatial extent. The exact definition of the virial radius is a matter of convention, but generically it refers to the radius at which the average density of enclosed matter is some factor larger than a reference background density. Some of the most common definitions set this threshold at 200 times the critical density (R_{200c}), or 200 or 360 times the cosmic matter density (R_{200m}, R_{360m}).

as a "standard ruler" to measure the expansion history of the Universe. In fact, recent measurements of the BAO peak from a variety of redshift surveys [e.g. Blake et al., 2011, Beutler et al., 2011, Anderson et al., 2012, Busca et al., 2013] have been able to provide an independent verification of ΛCDM using only the CMB as an external dataset, and without the need to include any other astrophysical probes (e.g. supernovae Ia, galaxy clusters, etc.).

Unlike in the case of optical galaxy samples, similar studies of galaxies selected by their 21cm atomic hydrogen (HI) emission have historically been at a much less developed stage. Until recently, HI-selected galaxy samples have been just too small to give us a representative view of the gas-rich galactic population. The advent of multi-feed radio receivers, however, made possible the execution of blind HI surveys, covering volumes comparable to those of optical surveys. The first generation 21cm HI Parkes All Sky Survey (HIPASS; Meyer et al., 2004), surveyed $\approx 30\,000$ deg^2 of sky (including the entire southern celestial hemisphere) and detected $5\,317$ sources[3]. This dataset enabled the first reliable measurements of the number density of galaxies as a function of their HI mass (HI mass function) and as a function of their HI linewidth (velocity width function), as well as their basic clustering properties and mass-velocity scaling relations. The main drawback of the HIPASS survey was its limited sensitivity: For example, the median redshift of the HIPASS sample was only $cz_{med} \approx 2\,500$ km s^{-1}, meaning that most HIPASS studies had to be confined to the very local Universe and thus cosmic variance was a major source of error. The limited sensitivity of HIPASS also meant that their sample contained a small number of low HI-mass and low-width sources (e.g. no HIPASS galaxies have $M_{HI} < 10^7\ M_\odot$), limiting the survey's power to study the low-mass end of the

galaxy population.

The Arecibo Legacy Fast ALFA[4] (ALFALFA; Giovanelli et al., 2005) survey was designed to address these limitations, and sample the gas-rich galactic population over a cosmologically significantly volume. The ALFALFA survey is conducted with the Arecibo 305m radio telescope, and as of November 2012 has completed data acquisition over an area of $\approx 6\,300$ deg^2. Despite the fact that the nominal volume of the ALFALFA survey is smaller than that of HIPASS, ALFALFA's increased sensitivity will allow it to detect almost an order of magnitude more HI sources. In fact, with its latest public data release –which covers the first $\approx 2\,800$ deg^2– ALFALFA has already produced the largest HI-selected sample to date ("α.40" catalog; Haynes et al., 2011), with a median redshift of $cz_{med} \approx 7\,500$ km s^{-1}. The α.40 catalog contains already ≈ 40 high signal-to-noise detections with $M_{HI} < 10^7\ M_\odot$ (and ≈ 300 with $M_{HI} < 10^8\ M_\odot$), making the ALFALFA dataset a powerful tool for studying the properties of the lowest-mass field galaxies.

This dissertation presents the research I conducted while a graduate student at the Astronomy Department of Cornell University, which is mostly focused on the statistical analysis of the unique ALFALFA dataset. The main goal of the presented research has been to shed light on the properties of dark matter on small scales and to reveal the physical mechanisms responsible for the collective properties of present-day galaxies, by analyzing the richest HI-selected galaxy sample sample available to date.

[3]Quoted numbers include the northern extension of HIPASS [Wong et al., 2006].

[4]The Arecibo L-band Feed Array (ALFA) is a 7-feed receiver operating in the L-band (≈ 1420 MHz), installed at the Arecibo Observatory.

In Chapter 2 we present the measurement of the velocity width function (WF) of galaxies from the 40% ALFALFA survey, which we compare with the expected distribution in a ΛCDM Universe. We report on the observed order-of-magnitude discrepancy at low velocity widths, and consider possible solutions related to the small-scale properties of dark matter and the internal kinematics of dwarf galaxies. In Chapter 3 we present the measurement of the baryonic mass function (BMF) of galaxies, i.e. the number density of galaxies as a function of their stellar and atomic gas mass, from ALFALFA and SDSS data. We combine the measured BMF with the ΛCDM halo mass function (HMF) through abundance matching, to show that low-mass halos are heavily baryon depleted even when their dominant atomic gas component is taken into account. Chapter 4 presents a detailed analysis of the clustering properties of the HI-selected ALFALFA sample. In particular, we study the dependence of clustering on galactic HI mass and we compare the clustering properties of HI-selected galaxies to those of optically selected samples from the SDSS. In addition, we use the clustering of ALFALFA galaxies to gain insight about the properties of halos hosting gas-rich galaxies. In Chapter 5 we summarize our main conclusions, and in Chapter 6 we end by discussing the future potential of similar studies in addressing open questions in galaxy formation and evolution, as well as in "near-field" cosmology.

In the Appendix we provide a detailed description of the "$1/V_{eff}$" statistical method used to measure the velocity width and baryonic mass function of galaxies, presented in Chapters 2 & 3. All distance-dependent quantities throughout this manuscript assume a present-day Hubble constant of $H_0 = 70$ $\mathrm{km\,s^{-1}\,Mpc^{-1}}$, unless otherwise specified.

CHAPTER 2

THE VELOCITY WIDTH FUNCTION OF GALAXIES FROM THE 40% ALFALFA SURVEY: SHEDDING LIGHT ON THE COLD DARK MATTER OVERABUNDANCE PROBLEM

2.1 Introduction

The current "standard" ΛCDM cosmological model has been extremely successful at reproducing the bulk of the observed properties of our universe on large scales [Komatsu et al., 2011]. However, given the current lack of a firm theoretical understanding of dark energy and the lack of a direct or indirect detection of the dark matter (DM) particle [Ahmed et al., 2009, Angle et al., 2008a,b, Abdo et al., 2010a,b, Adriani et al., 2009], it is important to test in detail the astrophysical implications of the established cosmological paradigm.

One of the most interesting consequences of assuming a cold dark matter (CDM) model is that substructure forms first on small scales, resulting in a present-day universe populated by a multitude of low-mass halos. More formally, the mass distribution of DM halos is described by the halo mass function (HMF), which is defined as the number density of halos as a function of their virial mass; it can be analytically predicted [Press and Schechter, 1974, Sheth and Tormen, 2002] that the HMF displays a power-law behavior at low halo masses, $n \propto M^{\alpha}$, with a relatively steep exponent of $\alpha \approx -1.9$ in the standard CDM context. This analytical expectation, confirmed to great accuracy by N-

*This chapter is an adapted version of the published article Papastergis et al. (2011).

body simulations of structure formation [Warren et al., 2006, Boylan-Kolchin et al., 2009, Klypin et al., 2010], leads to the prediction of a large number of low mass halos for every Milky Way-sized (MW-sized) halo found in the present epoch.

This firmly established theoretical result has led to a number of observational challenges, such as the "missing satellites problem" [Klypin et al., 1999, Moore et al., 1999, Diemand et al., 2007, Strigari et al., 2007, Simon and Geha, 2007], the "void phenomenon" [Peebles, 2001, Tinker and Conroy, 2009], as well as the discrepancy between the sizes of mini-voids observed in the local universe and those produced in CDM simulations [Tikhonov and Klypin, 2009]. Additional concerns, again closely related to the distribution of halo masses predicted by CDM, are raised by the flatness of the galactic luminosity function [LF, Blanton et al., 2005a, Montero-Dorta and Prada, 2009], HI mass function [HIMF, Martin et al., 2010] and galactic stellar mass function [SMF, Baldry et al., 2008, Li and White, 2009] at their faint/low-mass end. These observational distributions display power-laws with $\alpha \approx -1.3$, much shallower than expected from the combination of a CDM universe plus a naive linear relationship between halo mass and luminosity/baryonic mass. Despite their apparent diversity, all statements described above are just different aspects of the same fundamental issue: CDM structure formation predicts large numbers of low mass halos, seemingly in contradiction with the relative paucity of visible low-mass galaxies. Hereafter, we refer to this discrepancy as the *CDM overabundance problem*[1].

The main caveat regards the proper interpretation of these observational results. All phenomena mentioned so far rely on the measurement of quanti-

[1]This statement does not aim at including a second class of potential observational challenges to CDM, related to the density profile of halos in their central regions (known as the "cusp versus core" problem).

ties indirectly related to the mass of the hosting DM halo (e.g. luminosity or HI/stellar mass) and, as a result, do not provide a direct means of comparing the HMF expected for CDM with the HMF realized in nature. In fact, a number of environmental and feedback effects (see §2.5.2) are expected to affect the baryonic content of halos, with low mass ones being the most impacted.

Ideally, one would need a large sample of galaxies with directly measured dynamical masses (e.g. through lensing or satellite kinematics), extending all the way to the low mass regime. Unfortunately, current datasets are restricted to relatively massive galaxies. The best practical alternative would consist of a rich sample of resolved HI-interferometric rotation curves of galaxies, spanning a wide range in dynamical mass. Atomic hydrogen is usually the most spatially extended baryonic component in a galaxy, and therefore the best tracer of the rotation curve at large galactic radii. Such a sample could be used to determine the space density of galaxies as a function of their measured maximum rotational velocity, v_{rot}. This observational statistic, which is referred to as the velocity function (VF) of galaxies, is more directly related to the halo dynamics than statistics based on luminosity/baryonic mass and has a largely different set of systematic issues. However, current datasets are very limited, mostly because HI interferometry is time consuming (especially for low HI-mass targets), and requires large telescope arrays.

A more economical approach is to rely on wide-area, single-dish 21 cm surveys. Thanks to their intrinsic spectroscopic nature, HI surveys automatically obtain the spectral HI-line profile of every detected source. The velocity width of each detected galaxy, w, can thus be readily extracted, and the associated dataset can be used to measure the velocity width function (WF) of galaxies.

One can furthermore apply inclination corrections to the measured widths in order to retrieve intrinsic rotational velocities (v_{rot}), and then estimate the galactic VF. Correcting for inclination requires however the use of external datasets, usually optical/NIR photometric surveys.

Before this work, the most accurate WF and VF for late-type galaxies have been based on $4\,315$ and $2\,646$ HI-selected galaxies respectively, detected by the HIPASS survey [Zwaan et al., 2010, hereafter Zw10]. Their measurement of the VF extends over the velocity range $30\ \mathrm{km\,s^{-1}} < v_{rot} < 300\ \mathrm{km\,s^{-1}}$, and suggests a dramatic departure from the CDM expectation at low velocities ($v_{rot} \lesssim 100\ \mathrm{km\,s^{-1}}$). Recent determinations of the VF for massive early-type galaxies (which are mostly absent from HI-selected samples) have been obtained using the Sloan Digital Sky Survey (SDSS) and Two-degree Field Galaxy Redshift Survey (2dFGRS) datasets by Choi et al. [2007] and Chae [2010]. Both the late-type and early-type distributions are needed in order to derive the "total" galactic VF , since massive early-type galaxies are the dominant population at high velocities ($250\ \mathrm{km\,s^{-1}} \lesssim v_{rot} \lesssim 450\ \mathrm{km\,s^{-1}}$) while late-types dominate the counts at lower velocities ($v_{rot} \lesssim 250\ \mathrm{km\,s^{-1}}$).

In this chapter we present the Arecibo Legacy Fast ALFA (ALFALFA) measurement of the velocity width function of HI-bearing galaxies. The decision not to correct the measured widths for inclination is intentional, as the WF maintains all the advantages of the VF as a probe of the halo mass distribution, while featuring a number of observational advantages over the latter (see Sec. 2.3 for more details). The ALFALFA WF is based on $10\,744$ HI-selected galaxies (a more than twofold increase over previous datasets) and extends to widths as low as $w = 20\ \mathrm{km\,s^{-1}}$.

This chapter is organized as follows: in Section 2.2 we present the ALFALFA survey and the associated dataset; in Section 2.3 we discuss the observational advantages of the WF with respect to the inclination-corrected VF and present the ALFALFA measurement of the WF for HI-bearing galaxies; in Section 2.4 we address possible observational biases on the determination of the ALFALFA WF; in Section 2.5 we compare the ALFALFA measurement with the expectations in a CDM universe, and describe the possible solutions to the observed discrepancy at low widths; in Section 2.6 we derive the relation between v_{rot} (measured observationally) and v_{halo} (calculated from N-body simulations), that would be needed to reconcile the velocity distributions of CDM halos and observed galaxies. We conclude with Section 2.7 by summarizing our results.

We remind the reader that throughout this chapter we use a Hubble constant of $H_0 = 70\ \mathrm{km\,s^{-1}\,Mpc^{-1}}$; h_{70} refers to the Hubble constant in units of 70 $\mathrm{km\,s^{-1}\,Mpc^{-1}}$, while h refers to the Hubble constant in units of 100 $\mathrm{km\,s^{-1}\,Mpc^{-1}}$.

2.2 ALFALFA dataset

2.2.1 The survey

The ALFALFA survey is a wide-area, blind 21 cm emission-line survey that takes advantage of the increased survey speed offered by the 7-feed Arecibo L-band Feed Array (ALFA) receiver at the Arecibo Observatory. The ALFALFA data are acquired in a minimally invasive drift-scan mode in two passes, ideally separated by several months in order to enable the discrimination between narrow-band radio frequency interference (RFI) and small spectral width cos-

mic signals. As of November 2012, the survey has completed data acquisition over $\approx 6\,300$ deg^2 of sky and out to $cz \approx 18\,000$ km s^{-1}. Upon completion of data reduction and source extraction, ALFALFA is projected to detect $> 30\,000$ extragalactic sources. ALFALFA is more sensitive than the previous generation HIPASS survey [Meyer et al., 2004, Zwaan et al., 2004], with a 5σ detection limit of 0.72 Jy km s^{-1} for a source with a profile width of 200 km s^{-1} as compared to a 5σ sensitivity of 5.6 Jy km s^{-1} for the same source in HIPASS [Giovanelli et al., 2005]. In addition to greater sensitivity, ALFALFA has a finer velocity resolution (11.2 km s^{-1} versus 26.4 km s^{-1} for smoothed data) and better angular resolution ($3.6'$ vs. $13'$ FWHM), resulting in a more accurate identification of optical counterparts.

2.2.2 The sample

The work presented in this chapter makes use of the ALFALFA data sample available prior to December 2010. In particular, catalogs had up to then been extracted for a total area of $2\,934$ deg^2 [Giovanelli et al., 2007, Saintonge et al., 2008, Kent et al., 2008, Stierwalt et al., 2009, Martin et al., 2009, Haynes et al., 2011]. At that time, the ALFALFA footprint consisted of four distinct regions: two in the northern Galactic hemisphere, referred to as the Virgo direction region for the remainder of this chapter (VdR: $07^h30^m < \alpha < 16^h30^m$, $4^\circ < \delta < 16^\circ$ and $24^\circ < \delta < 28^\circ$), and two in the southern Galactic hemisphere, referred to as the anti-Virgo direction region (aVdR: $22^h < \alpha < 03^h,\ 14^\circ < \delta < 16^\circ$ and $24^\circ < \delta < 32^\circ$). From this primary dataset we only selected extragalactic objects detected at high significance ($S/N > 6.5$, designated "Code 1"), and we further restricted ourselves to the redshift range $cz \leqslant 15\,000$ km s^{-1}, beyond which

interference from the nearby San Juan airport causes a significant drop of the ALFALFA detection efficiency. The final sample used in this work, corresponding to slightly less than 40% of the ALFALFA survey area (hereafter "$\alpha.40^{-}$" sample), contains thus a total of $11\,086$ galaxies.

Figure 2.1 shows the spatial distribution of the $\alpha.40^{-}$ sources in the Virgo and anti-Virgo directions respectively, and puts in evidence the complex large-scale structure present in both volumes. Density fluctuations in the survey volume can be the dominant source of statistical uncertainty in surveys like ALFALFA, where the large sample size ensures small counting errors. Our statistical estimator, described in §2.3.1, in chosen to minimize this structure-induced bias.

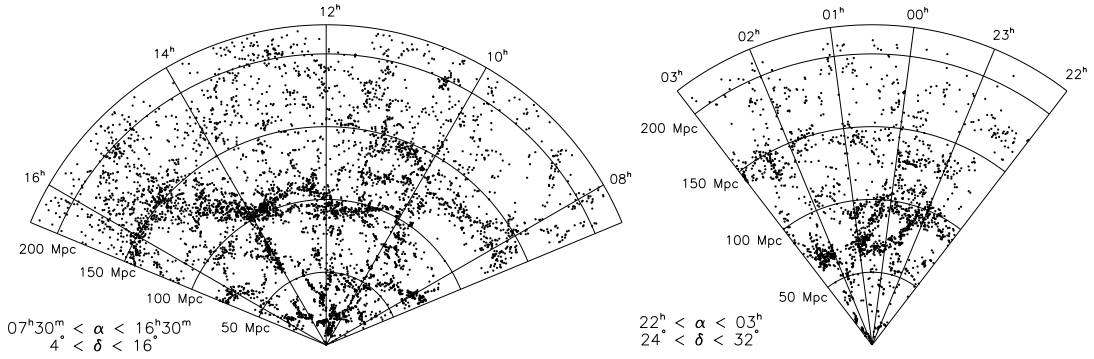

Figure 2.1 Spatial distribution of $5\,868$ sources in the Virgo direction region (VdR, left panel) and $2\,055$ sources in the anti-Virgo direction region (aVdR, right panel). The Virgo Cluster and the "Great Wall" are the most conspicuous structures in the VdR (located at a distance of ≈ 17 Mpc and ≈ 100 Mpc respectively). In the aVdR, the Pisces-Perseus Supercluster (clearly visible at ≈ 70 Mpc) as well as the void in its foreground dominate the large-scale structure. Distances are assigned through a combination of a flow model for the nearby universe [Masters, 2005] and Hubble distances for more distant galaxies (see §2.2.2).

Figure 2.2 displays some statistical properties of the $\alpha.40^{-}$ sample. Histograms (a) and (b) represent the distribution of heliocentric velocity, $v_{\odot}$, and of signal profile width, w_{50}, which are both directly measured quantities [Giovanelli et al., 2007, Saintonge, 2007]. The signal profile width is measured at the

50% flux level of each of the two peaks of the typical double-horned HI profile (or at 50% of the single peak flux, for single-peaked profiles). The value of w_{50} reported in the ALFALFA catalogs is further corrected for instrumental broadening. Histogram (c) displays the distribution of galaxy HI mass, M_{HI}, which is a distance dependent (and hence derived) quantity. Unlike previous HI surveys, we assign distances to nearby galaxies through a peculiar velocity flow model of Masters [2005], and use Hubble flow distances only for galaxies with $cz > 6\,000\ \mathrm{km\,s^{-1}}$ (see §2.4.2 for a detailed discussion on the impact of distance uncertainties on ALFALFA results).

Figure 2.3 displays the distribution of $\alpha.40^{-}$ sources in the velocity width (w_{50}) versus integrated-flux (S_{int}) plane. As expected, the detection limit of the survey is a function of signal profile width and correctly scales as $S_{int,lim} \sim w_{50}^{1/2}$. Due to the large density of sources near the detection limit, we evaluate the completeness limit of the survey (red dashed line in Figure 2.3) based on the actual data rather than on simulations using synthetic sources.

2.3 The Velocity Width Function

We obtain rest-frame galaxy velocity widths, w, by correcting the cataloged profile widths (w_{50}) for Doppler broadening. It is customary to apply additional inclination corrections to w, in order to recover intrinsic rotational velocities, v_{rot}. However, since most extragalactic sources are unresolved at centimeter wavelengths, such corrections rely on external datasets (usually optical or NIR photometric surveys) for the determination of galaxy inclinations. Here, we choose to make no further corrections to w and measure the velocity width function

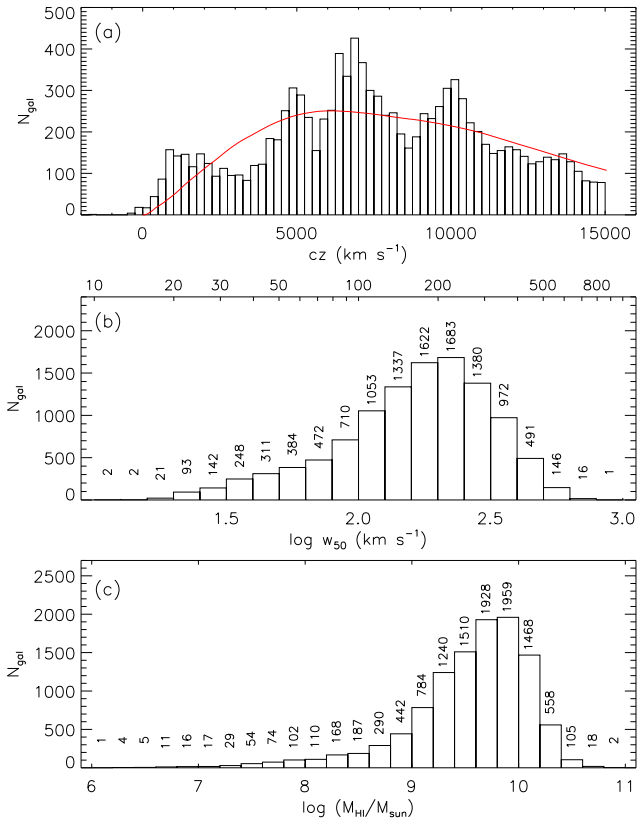

Figure 2.2 Histogram (a) represents the distribution in heliocentric velocity ($v_\odot$), while the red solid line represents the distribution expected in a homogeneous universe according to the selection function of the survey; the complex large-scale structure in the survey volume is apparent. Histogram (b) represents the distribution of velocity width (w_{50}); note the large number of very low-width galaxies ($w_{50} < 30$ km s^{-1}) detected. Histogram (c) represents the distribution of galaxy HI mass (M_{HI}); again note the detections at very low HI mass ($M_{HI} < 10^8\ M_\odot$).

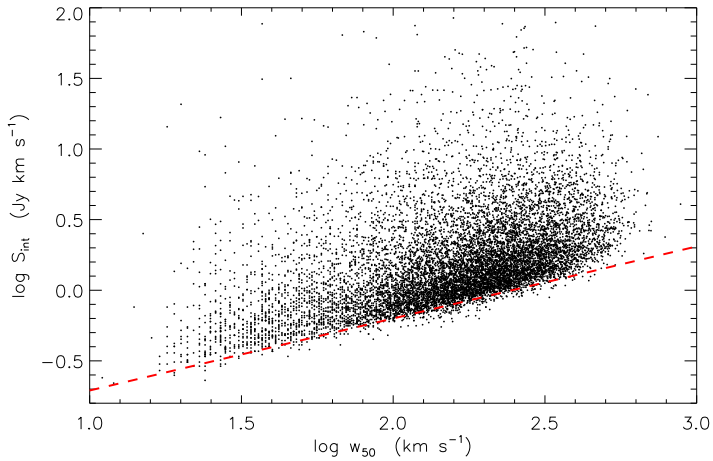

Figure 2.3 Distribution of the $\alpha.40^{-}$ sources in the velocity width vs. integrated flux ($w_{50} - S_{int}$) plane. The dashed red line is the survey completeness limit adopted in this work $(S_{int,lim}/1 \text{ Jy km s}^{-1}) = 0.06\ (w_{50}/1 \text{ km s}^{-1})^{0.51}$, which follows very closely the theoretically expected $S_{int,lim} \propto w^{1/2}$.

(WF) of galaxies, denoted by $\phi(w)$.

Even though the WF does not directly represent the distribution of any fundamental galaxy property, it is observationally superior to the (inclination-corrected) VF. In particular, it is free of the restrictions and systematics that arise from cross-matching HI and optical catalogs and correcting for galaxy inclination. For example, the HIPASS primary sample contains $4\,315$ sources of which only $2\,646$ have unambiguous optical counterparts [Doyle et al., 2005]. Another 30% of the sources in this restricted subsample have low inclination values ($i < 45°$), and are thus subject to large inclination-correction errors. As a result, only $\approx 43\%$ of the galaxies in the HIPASS primary sample were used for their determination of the VF. Furthermore, obtaining accurate estimates of the true orientation of irregularly-shaped dwarf galaxies is challenging, and the process may introduce biases in the measurement of the low-velocity end of the VF.

Nonetheless, a measurement of the galactic WF would not be useful if it did not provide an accurate means of comparing the outcome of N-body simulations with the observed universe. Fortunately, it is relatively straightforward to project a given theoretical rotational velocity distribution and transform it into its corresponding width distribution (see §2.5.1). We conclude that the WF should be regarded as the *prime observational distribution* for single-dish HI surveys, against which to compare theoretical expectations.

2.3.1 The ALFALFA Velocity Width Function

In Figure 2.4 we present the ALFALFA width function, based on $10\,744$ galaxies drawn from the $\alpha.40^{-}$ sample. For the calculation of the WF we restrict ourselves to $\alpha.40^{-}$ galaxies which are positioned in the portion of the flux-width plane where the ALFALFA survey is complete (i.e. above the red dashed line in Figure 2.3) and have profile widths broader than $w_{50} \gtrsim 18\ \mathrm{km\,s^{-1}}$. This cut results in the elimination of ≈ 330 galaxies from the calculation. An additional 13 very nearby sources are eliminated, for which the flow model assigned distances are subject to large uncertainty.

The WF is calculated in logarithmic width bins, according to the $\Sigma\, 1/V_{eff}$ method [Zwaan et al., 2005]. The $\Sigma\, 1/V_{eff}$ method is a non-parametric maximum likelihood method and, as such, it is insensitive to the presence of large-scale structure in the survey volume. As its name suggests, it closely resembles the traditional $\Sigma\, 1/V_{max}$ method [Schmidt, 1968] and consists of summing the number of detections in each width bin, weighted by the inverse of the "effective" volume available to each source. More precisely, the space density of

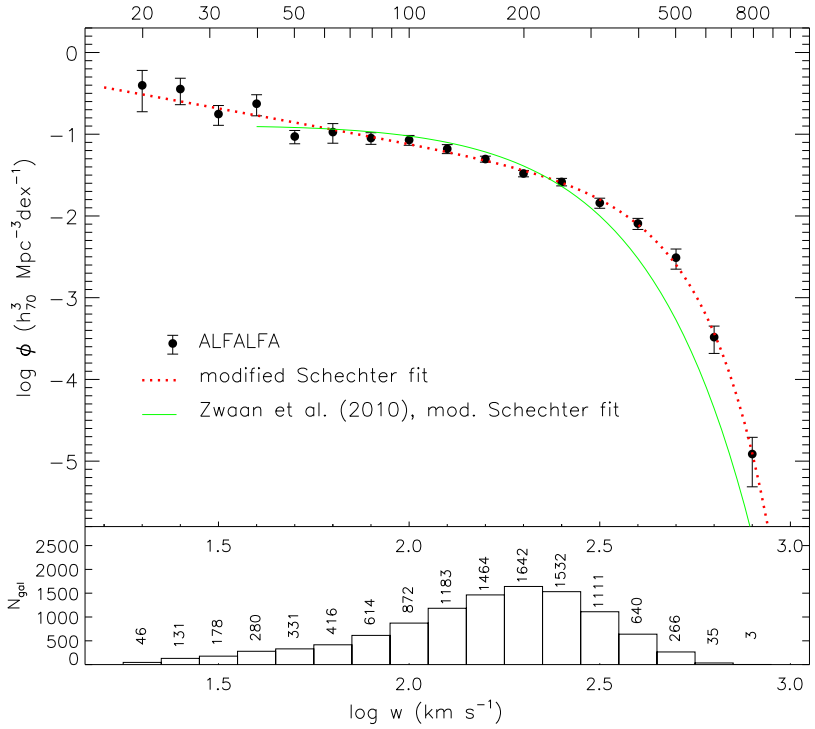

Figure 2.4 Datapoints represent the space density of HI-bearing galaxies as a function of velocity width (corrected for Doppler and instrumental broadening, but uncorrected for inclination), as inferred from 10 744 galaxies detected by the 40% ALFALFA survey. The errors are 1σ Poisson errors due to galaxy counts in individual width bins. The red dotted line corresponds to a modified Schechter fit to the ALFALFA WF (see §2.3.1). The green solid line represents the fit to the HIPASS WF based on 4 315 galaxies [Zwaan et al., 2010], over its measured range.

galaxies belonging to width bin k ($k = 1, 2, ..., N_w$) is

$$\phi_k = \sum_i \frac{1}{V_{eff,i}} \quad \text{for all galaxies } i \text{ in width bin } k \ . \tag{2.1}$$

In the case of a spatially homogeneous survey volume, $V_{eff,i}$ would coincide with $V_{max,i}$, the latter defined as the volume within which galaxy i could be placed and still be detectable by the survey. On the other hand, if the survey volume displays significant density variations, $V_{eff,i}$ takes into account the relative density of the volume available to galaxy i with respect to the mean density of the total survey volume. As with all density-independent estimators, the overall normalization is lost, and has to be calculated afterwards. The normalization is fixed by matching the integral of the distribution to the average number density of galaxies in the survey volume (a full description of the method can be found in the Appendix).

Due to its spectral resolution and sensitivity, ALFALFA can push the low-width limit of the WF to $w \approx 20\ \mathrm{km\,s^{-1}}$, a factor of 2 lower than the HIPASS survey. Over the full measured range ($20\ \mathrm{km\,s^{-1}} < w < 800\ \mathrm{km\,s^{-1}}$) the ALFALFA WF is very well described by a modified Schechter function of the form[2]

$$\phi(w) = \frac{dn}{d\log w} = \ln(10)\ \phi_* \left(\frac{w}{w_*}\right)^{\alpha} e^{-(\frac{w}{w_*})^{\beta}} \ . \tag{2.2}$$

The least squares parameters[3] are $\phi_* = 0.011 \pm 0.002 \quad h_{70}^3$ Mpc^{-3}dex^{-1}, $\log w_* = 2.58 \pm 0.03$, $\alpha = -0.85 \pm 0.10$ and $\beta = 2.7 \pm 0.3$ (uncertainties are statistical 1σ errors due to Poisson errors on the individual bin values). Note,

[2]The parameterization used here is equivalent to the parameterization $\phi(w)\ dw = \phi_*\ (w/w_*)^{\alpha}\ \exp -(w/w_*)^{\beta}\ (\beta/\Gamma(\alpha/\beta))\ dw/w$ presented by other authors, except for the normalization factor $\beta/\Gamma(\alpha/\beta)$.

however, that the final sample contains 163 sources that lack a confidently identified optical counterpart. Some of these sources correspond to tidal debris from nearby interacting galaxies and may not be hosted by individual DM halos. Excluding these galaxies from the WF calculation leads to a somewhat shallower narrow-end slope of $\alpha = -0.68 \pm 0.11$.

ALFALFA finds significantly more high-width galaxies than HIPASS (a factor of ~ 3 at $w \approx 400$ km s^{-1}, growing to a factor of ~ 10 at $w \approx 800$ km s^{-1}), which is also evident from the marked difference in the value of the position of the "knee" of the WF for the two surveys ($\log w_* = 2.58 \pm 0.03$ for ALFALFA versus $\log w_* = 2.21 \pm 0.10$ for HIPASS[4], in disagreement at the $> 3\sigma$ level). Despite the fact that the nominal HIPASS volume is a factor of ~ 5 larger than the $\alpha.40^-$ volume, ALFALFA is able to find more high-width galaxies thanks to its better sensitivity (see Figure 2.5). The same effect can be seen in the HIMFs published by the two surveys, with ALFALFA [Martin et al., 2010] finding a factor of a few more of the highest HI-mass galaxies compared to HIPASS [Zwaan et al., 2005].

On the low-width end, ALFALFA finds a rising slope ($\alpha < 0$) which is, however, by no means steep enough to match the CDM prediction (see Sec. 2.5). Despite the vastly different value for the narrow-end slope reported by the two surveys ($\alpha = 0.10 \pm 0.39$ for HIPASS versus $\alpha = -0.85 \pm 0.10$ for ALFALFA) the HIPASS and ALFALFA datapoints are consistent in the width range 40 km s^{-1} $\lesssim w \lesssim 200$ km s^{-1}. The HIPASS α parameter is not well constrained, as their WF does not extend to low enough widths and suffers from considerable counting

[3]The least squares parameters and their statistical errors were determined by the MPFITFUN procedure, written in the IDL programming language.

[4]No errors are reported for the published fit parameters to the HIPASS WF. In order to compare with ALFALFA, we derive errors by performing a least squares fit to the HIPASS WF datapoints.

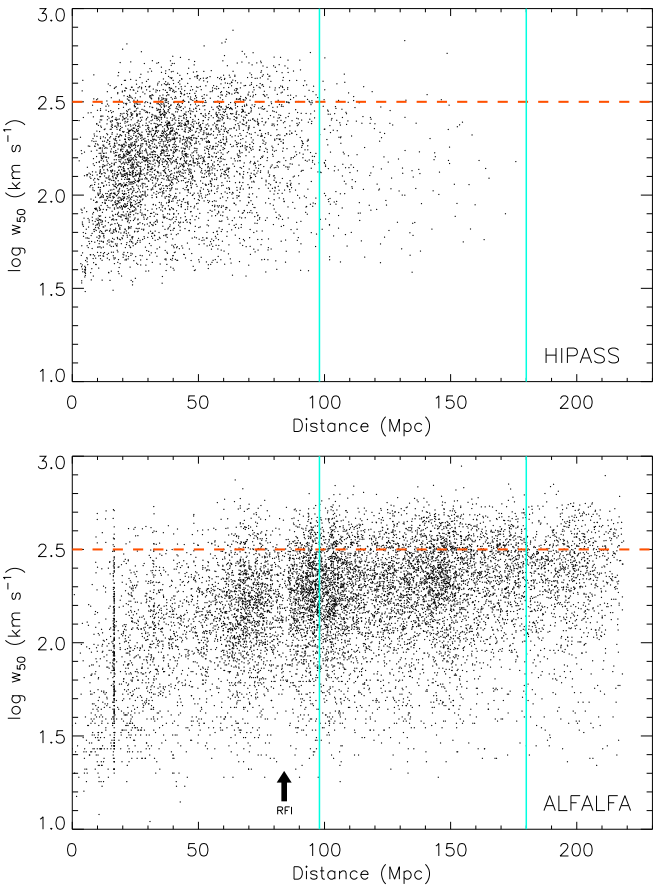

Figure 2.5 The figure shows the "velocity width Spanhauer" diagrams for ALFALFA (lower panel) and HIPASS (upper panel) on the same scale. The region above the horizontal orange line marks the range over which the two width functions disagree. Despite the fact that the nominal value of the HIPASS volume is a factor of $\sim$5 larger than the α.40 volume, ALFALFA detects more very broad profile galaxies. This is due to the limited sensitivity of HIPASS, which leads to a "thinning" of detections beyond $\approx$ 100 Mpc and out to the survey boundary (area enclosed by vertical cyan lines).

error in the low-width bins.

2.4 Biases

2.4.1 Linewidth measurement errors

Measurement errors on w_{50} can shift galaxies among width bins, altering the bin counts and therefore the inferred space density. The w_{50} value for ALFALFA sources is subject to two separate sources of error: one is statistical in nature and present for all sources, while the other is systematic and concerns only a fraction of the $\alpha.40^{-}$ sample. The former is due to the distortion of the signal profile shape by noise; the latter results from the fact that the measurement of the spectral width of a signal relies on the accurate visual identification of its spectral boundaries, which is non-trivial for a number of sources (especially those found in the vicinity of RFI). The final width error reported in the ALFALFA catalogs, Δw_{50}, is the sum in quadrature of the random and systematic error terms described above. Owing to the fact that all $\alpha.40^{-}$ galaxies are detected with high signal to noise and have a clean spectral profile in the vast majority of cases, the typical $\alpha.40^{-}$ width error is relatively small and its distribution well behaved. The median error is $\Delta w_{50,median} \approx 8\ \mathrm{km\,s^{-1}}$ and $\sim$70% of the sources have a fractional error of $\Delta w_{50}/w_{50} \leqslant 10\%$.

In order to assess the effect of Δw_{50} on the WF, we create 50 mock galaxy samples by re-assigning widths to every galaxy i in the primary ALFALFA dataset according to their individual measured width ($w_{50,i}$) and error ($\Delta w_{50,i}$). Each mock sample is subject to the same cuts as the $\alpha.40^{-}$ sample and a new real-

ization of the WF is calculated ("1x" set). In order to illustrate the systematic trends introduced, we also perform an additional set of WF realizations with artificially inflated width errors (twice the reported ALFALFA width errors, "2x" set).

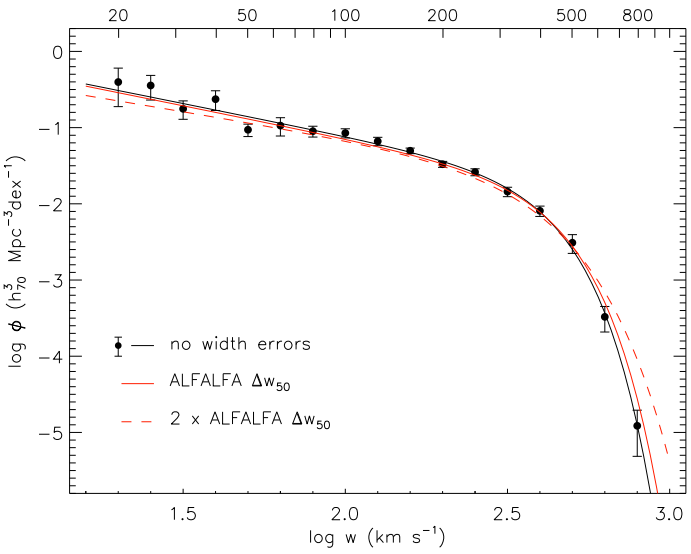

Figure 2.6 Filled circles with errorbars and the black solid line represent the ALFALFA WF and the best-fitting modified Schechter function (same as Figure 2.4). The red solid line corresponds to the distribution obtained by taking into account the ALFALFA measurement error on w_{50}. The WF remains mostly unchanged, except perhaps for a slight increase at the high width end. The red dashed line corresponds to artificially inflated width errors (twice the α.40 errors) and is plotted in order to illustrate the general systematic trend introduced by width errors on the WF (see §2.4.1 for more details).

The results are shown in Figure 2.6: overplotted to the original ALFALFA WF (datapoints and solid black line) are a modified Schechter fit to the mean WF corresponding to the 1x (red solid line) and 2x (dashed red line) realizations. As evidenced by the 2x run, width errors generally lead to a rise of the high-width end of the WF, due to a net "diffusion" of galaxies from intermediate-width bins with large number counts to high-width bins with low number counts. However, width errors at the ALFALFA error levels (1x set) seem to have a relatively minor effect on the WF.

2.4.2 Distance Uncertainties

Since velocity width is a distance-independent quantity, galaxy counts in width bins are not altered by distance errors. However, the weights ($1/V_{eff,i}$) that each galaxy contributes to its bin depend on HI-mass (see Eqn. A.14 and discussion in §2.3.1), and therefore on the assumed distance. Masters et al. [2004] have shown that ignoring the local peculiar velocity field can lead to biased estimates of galaxy statistical distributions, especially for surveys drawing a large fraction of their sample from the Virgo direction (VdR). To avoid this bias ALFALFA uses redshift distances only for distant ($cz > 6\,000$ km s^{-1}) galaxies and assigns distances to nearby galaxies through a parametric flow model developed by Masters [2005]. The model includes two attractors (Virgo Cluster & Great Attractor), a dipole component (Local Group peculiar velocity), a quadrupole component (Local Group asymmetric expansion) and a random thermal residual of $\sigma_{local} \approx 160$ km s^{-1}. Here we assume that most of the coherent motion of nearby galaxies is correctly described by the flow model, and no significant bias results from this systematic component of galaxy peculiar velocities. Contrary to intuition however, even the random component σ_{local} can induce a systematic bias through the "Eddington effect" [see for example Figure 6 in Zwaan et al., 2003].

In order to asses the effect of σ_{local} on the WF, we proceed as in §2.4.1 and create 50 mock samples by adding Gaussian noise on the cataloged distance of each $\alpha.40^{-}$ galaxy. We calculate the WF corresponding to each sample realization, and use the obtained average distribution to investigate the effect of distance uncertainties on the WF. We adopt the Masters [2005] value of $\sigma_{local} \approx 160$ km s^{-1} for our distance noise, but we also perform simulations with double the fiducial dispersion ($\sigma_{local} \approx 320$ km s^{-1}).

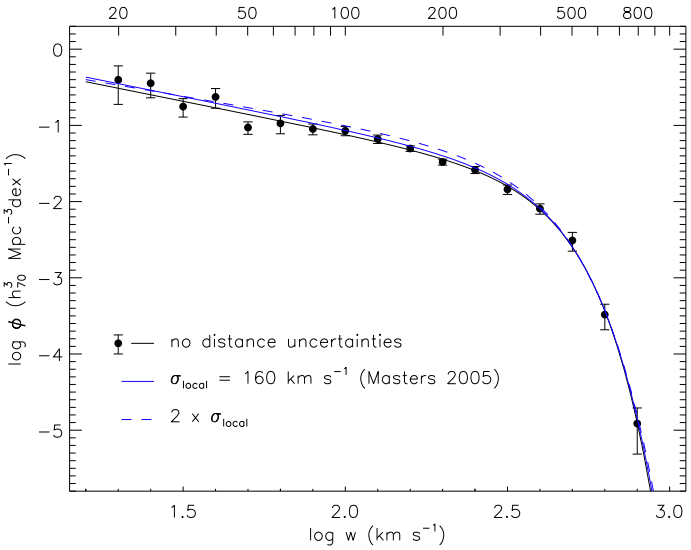

Figure 2.7 Filled circles with error bars and black solid line as in Figure 2.4. The blue solid line corresponds to the result of adding a random velocity dispersion of $\sigma_{local} = 160\ \mathrm{km\,s^{-1}}$ [Masters, 2005] to the α.40 galaxy distances. The dashed blue line corresponds to twice the fiducial velocity dispersion, $\sigma_{local} = 320\ \mathrm{km\,s^{-1}}$. Note the relative immunity of the WF against distance uncertainties. The main effect appears to be an overall increase in amplitude, while (in contrast to the case of the HIMF) no clear trend for a steepening of the low-end slope seems to exist (see §2.4.2 for more details).

The results are displayed graphically in Figure 2.7, where the datapoints and solid line correspond to the original ALFALFA WF, and the blue solid and dotted lines correspond respectively to the results of the $\sigma_{local} = 160 \text{ km s}^{-1}$ and $\sigma_{local} = 320 \text{ km s}^{-1}$ simulation sets. The largest effect is an overall increase in the amplitude of the WF. This is probably due to the fact that distance errors will preferentially move sources to distances smaller than their true value; this will bias a number of sources towards lower HI masses, which then translates into larger values for their $1/V_{eff}$ weights. On the other hand, and unlike in the case of the HIMF, the low end slope α does not seem to be heavily affected. We conclude that, apart from a mild increase in amplitude at intermediate widths, the WF is relatively insensitive to distance uncertainties due to galaxy peculiar motions.

2.4.3 Cosmic Variance

The WF presented in Figure 2.4 aspires to represent the distribution in a cosmologically representative volume. The sensitivity of ALFALFA allows $\sim w_*$ and broader galaxies to be detected throughout the full survey volume ($V_{survey} \approx 3.1 \cdot 10^6 \ h_{70}^3 \text{ Mpc}^3$), which ensures a cosmologically fair sampling of the MW-sized galaxy population. On the other hand, low-width galaxies tend to be faint systems that can only be detected locally. As a result, the low-width bins of the WF are subject to increased uncertainty caused by the deviation of the galaxy distribution from homogeneity on small scales, which is referred to as cosmic variance (see Figure 2.8 for a graphical illustration).

In order to quantitatively asses the effects of cosmic variance on the AL-

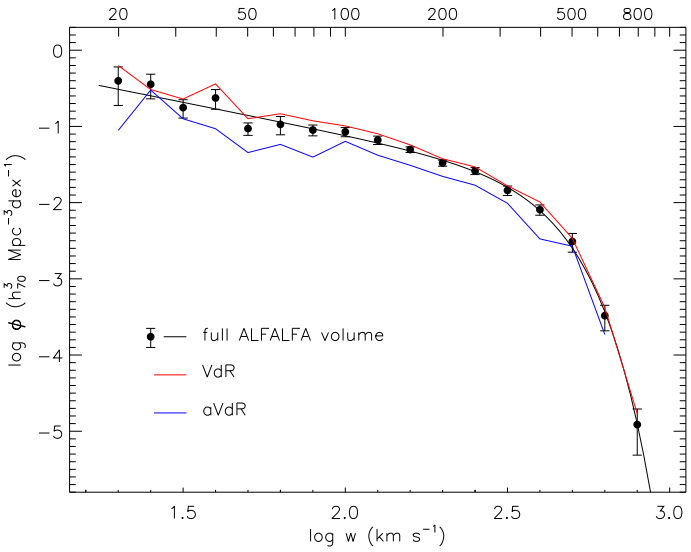

Figure 2.8 Datapoints with error bars and black solid line as in Figure 2.4. The red and blue solid lines represent the WF in the Virgo direction Region (VdR: $07^h30^m < \alpha < 16^h30^m$, $4° < \delta < 16°$ and $24° < \delta < 28°$) and the anti-Virgo direction Region (aVdR: $22^h < \alpha < 03^h$, $14° < \delta < 16°$ and $24° < \delta < 32°$), respectively. The VdR is a locally overdense region while the aVdR is locally underdense, a fact that is reflected by the difference between the the two WFs at intermediate and low widths (see §2.4.3).

FALFA WF, we jackknife resample the $\alpha.40^-$ sample, by splitting it into 14 parts equally spaced in R.A. Then, we reevaluate the WF excluding each part in turn. The resulting scatter for each fit parameter, x, is given by

$$\sigma_x^2 = \frac{N-1}{N} \sum_i (x_i - \overline{x})^2 \ , \quad i = 1, \ldots, 14. \tag{2.3}$$

The scatter calculated by Eqn. 2.3 would be equal to the purely statistical error if the survey volume were homogeneous, and so any excess noise results from the presence of inhomogeneities. The method described above provides a measurement of cosmic variance on linear scales smaller than those probed by the full survey, and hence yields a conservative estimate of the true uncertainty (cosmic variance generally increases with decreasing scale).

The full uncertainties on the fit parameters (including cosmic variance) are $\phi_* = 0.011 \pm 0.003\ (0.002)\ \ h_{70}^3$ Mpc^{-3}dex^{-1}, $\log w_* = 2.58 \pm 0.04\ (0.03)$, $\alpha = -0.85 \pm 0.19\ (0.10)$ and $\beta = 2.7 \pm 0.3\ (0.3)$, where the term in parentheses represents the purely Poisson error reported in §2.3.1. Indeed, parameters w_*, ϕ_* and β, which dictate the shape of the WF mostly at high widths, show a very modest increase in their uncertainty due to cosmic variance. On the other hand, the narrow end slope α is significantly affected, with cosmic variance contributing a large fraction of the full error.

2.4.4 Beam confusion

Beam confusion arises from the fact that the ALFA $3.3'$ x $3.8'$ beam occasionally produces blends of close pairs of galaxies or small galaxy groups, especially when individual galaxies are poorly separated in redshift space. The qualitative effect of such blends is to transform two or more independent sources into a single HI profile of larger w_{50} than each of its constituents. We do not attempt to quantify the effect of confusion bias, but we anticipate it to be more pronounced at the high-width end of the WF. This is because galaxies with large widths usually correspond to HI massive objects, which are rare and thus preferentially found at large distances.It is worth noting that this bias, even though present, cannot account for the discrepancy between the ALFALFA and HIPASS WFs at high widths, since the latter suffers from more confusion due to its larger beam size ($13'$ FWHM).

2.5 Comparison with Theory and Simulations

The velocity function of halos in a CDM universe scales as $dn \propto v^{-4}\, dv$, where v refers to the halo maximum rotational velocity. Even though a straightforward comparison of the CDM VF with the ALFALFA WF is not possible, such a steep scaling suggests a substantial discrepancy between the theoretical and observed distributions at low velocities.

In order to make a meaningful comparison between the theoretical prediction and the ALFALFA measurement, it is necessary to take into account a number of important effects:

1. In general, a one-to-one correspondence between simulated halos and visible galaxies is not always possible. For example, massive halos ($M_{halo} \gtrsim 10^{13}\ h^{-1}\, M_{\odot}$) typically host groups or clusters of galaxies, rather than a single astronomical object.
2. The collapse of baryons to the central region of DM halos affects the galactic potential and leads to a modification of the true galactic rotation curves compared to the ones obtained in dissipationless DM simulations.
3. The detectability of a galaxy in an HI survey depends on its atomic hydrogen content. Galaxies that are deficient in HI will be underrepresented in an HI-selected sample.
4. The relationship between the maximum of the rotation curve of a galaxy and its HI velocity width is non-trivial. Apart from the obvious dependence on disk inclination, the measured width depends on the spatial distribution of atomic hydrogen in the galactic potential. In particular, HI disks do not always extend far enough to sample the asymptotic outer

part of the galactic rotation curve, and may underestimate the mass of the host halo.

It is, thus, necessary to populate the DM halos of an N-body simulation with *modeled* galaxies, and compare this virtual sample against the ALFALFA measurement. Modeling of the atomic hydrogen content of the synthetic galaxies is particularly desirable, because it greatly facilitates the comparison between theoretical and observed distributions.

Obreschkow et al. [2009, hereafter O09] have simulated the HI-line profiles for the galaxies in the De Lucia and Blaizot [2007] semi-analytic catalog, created by post-processing the Millennium N-body simulation [Springel et al., 2005]. Figure 2.9 displays the WF (cyan solid line) resulting from projecting their modeled edge-on linewidths, assuming random galaxy inclinations. The O09 WF is in fairly good agreement with the ALFALFA measurement, but fails to display an exponential cutoff at high widths and therefore predicts too many high-width galaxies. This issue has been also pointed out in Zw10, who argue that the disagreement is caused by the fact that the O09 catalog overestimates the HI masses of massive early-type galaxies. They found that restricting themselves to synthetic galaxies classified as late-types (based on their bulge-to-total stellar mass ratios in the DeLucia catalog) produced a much better fit to their data. However, Figure 2.9 suggests that applying the "morphological" cut of Zw10 results in too few galaxies at intermediate widths ($200\,\mathrm{km\,s^{-1}} < w < 600\,\mathrm{km\,s^{-1}}$).

The red solid line in Figure 2.9 is the WF corresponding to an indirect observational estimate of the velocity distribution of spiral galaxies by Gonzalez et al. [2000]. Their determination of the spiral galaxy VF was produced by combining the Southern Sky Redshift Survey B-band LF for spirals in conjunction with

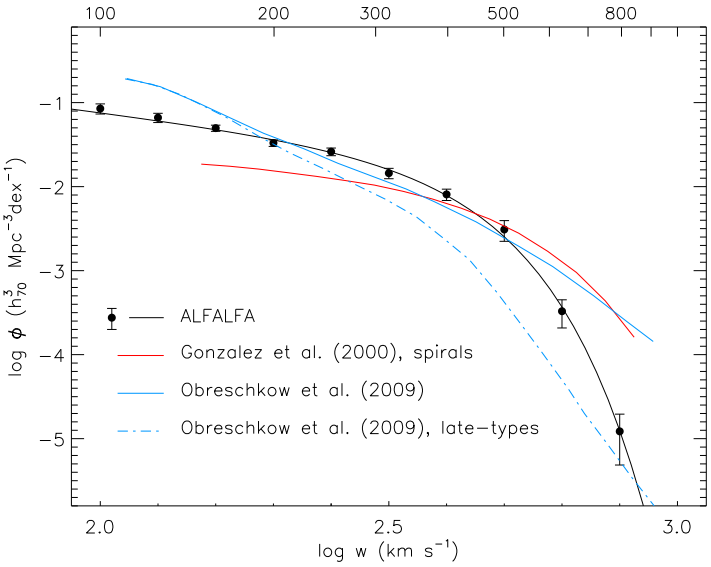

Figure 2.9 Datapoints with errorbars and the black solid line represent the ALFALFA WF in the width range $w \geqslant 100$ km s^{-1}. The cyan solid line represents the Obreschkow et al. [2009, O09] WF, derived from projecting their distribution of modeled HI linewidths (w_{50}^{HI}) for the synthetic galaxies in the semi-analytic catalog of De Lucia and Blaizot [2007]. The cyan dash-dotted line represents the subsample of the O09 galaxies classified as "late-types" according to their bulge-to-stellar mass ratios in the DeLucia catalog. The red solid line represents the projection of the indirect observational determination of the velocity function (VF) of spiral galaxies by Gonzalez et al. [2000]. Their VF was obtained by combining the observed luminosity function (LF) for spiral galaxies with the Tully-Fisher relationship.

the Yasuda et al. [1997] Tully-Fisher parameters in the B_T-band. This indirect method, based on galaxy scaling relations, is reliable only for relatively massive spirals ($v_{rot} > 70\ \mathrm{km\,s^{-1}}$) and suffers from numerous sources of uncertainty (e.g. scatter in the TF relation, uncertainties related to the correction of the LF for extinction, bandpass conversion uncertainties, etc.).

2.5.1 The CDM overabundance problem

CDM predictions start diverging from the observational results at low widths, and so the behavior of the theoretical WF for $w < 200\ \mathrm{km\,s^{-1}}$ is of great importance. Unfortunately, the very interesting work of O09 is only reliable for $w \gtrsim 100\ \mathrm{km\,s^{-1}}$ due to the limitations in the mass resolution of the Millennium simulation. To compare with the ALFALFA data, we employ instead two virtual galaxy samples populating recent high-resolution CDM simulations.

Figure 2.10 compares the ALFALFA measurement with the WF of the galaxy population corresponding to the Bolshoi simulation [Klypin et al., 2010], as modeled by Trujillo-Gomez et al. [2010, hereafter TG10]. Each Bolshoi halo was assigned realistic stellar and cold gas masses, based on empirical relations. Subsequently, two models were considered, one where the gravitational potential of the baryons is simply superimposed on the DM potential (solid green line) and one where the DM halo adiabatically contracts in response to the presence of the baryons (dash-dotted green line). Note that TG10 define v_{rot} as the value of the simulated rotation curve at a radius of 10 kpc. The authors argue that their modeling scheme and use of v_{10kpc} provide a good approximation of the measured velocity for galaxies with both flat and rising rotation curves.

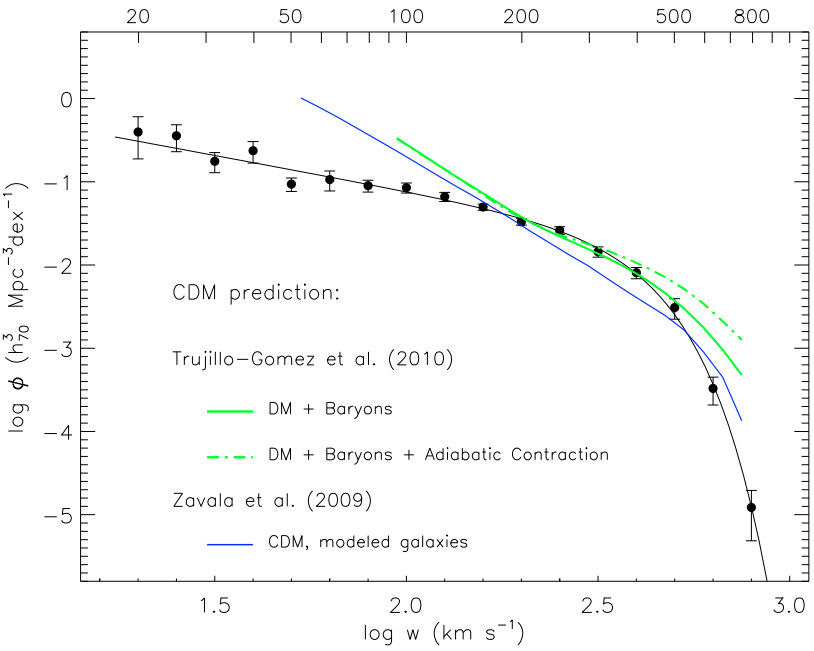

Figure 2.10 The CDM overabundance problem: datapoints with errorbars and black solid line represent the measured ALFALFA WF (same as in Figure 2.4). The green lines represent the WF of a sample of synthetic galaxies modeled by Trujillo-Gomez et al. [2010, TG10], which populate the halos in the Bolshoi CDM simulation [Klypin et al., 2010]. Two models were considered by TG10, one where the gravitational potential of baryons is simply superimposed on the DM potential (solid line) and one where the subsequent adiabatic contraction of the DM halo is taken into account (dash-dotted line). The blue solid line represents the WF of a modeled galaxy population corresponding to the higher resolution CDM simulation of Zavala et al. [2009, Za09]. Note that both theoretical distributions predict a steeply rising low-width end, in stark contrast with the observational result. The discrepancy according to the Za09 result is a factor of $\sim$8 at $w = 50$ km s^{-1}, rising to a factor of ~ 100 when extrapolated to $w = 20$ km s^{-1} (see §2.5.1).

Also plotted in Figure 2.10 is the WF of simulated galaxies based on the Zavala et al. [2009, hereafter Za09] constrained N-body simulation (blue solid line). Za09 produced a modest volume ($64\ h^{-1}$ Mpc on a side) but very high-resolution ($v_{lim} = 24\ \mathrm{km\,s^{-1}}$) constrained simulation, designed to reproduce the large-scale structure of the local universe. Virtual galaxies are modeled according to the analytical results of Mo et al. [1998], assuming a disk-to-virial mass ratio of $f_{disk} \equiv M_{disk}/M_{vir} = 0.03$ independent of halo size. Lastly, the maximum amplitude of the rotation curve ($v_{rot,max}$) for each galaxy is calculated, by combining the disk and DM halo contributions.

Note, however, that both simulation works presented here lack detailed modeling of the HI component of their virtual galaxy samples, which is the the velocity field tracer. As a result, we convert rotational velocities into HI velocity widths by the following analytic expression:

$$w = 2\, v_{rot}\, \sin i + w_{eff} \quad . \tag{2.4}$$

Galaxies are assumed to be randomly oriented with respect to the line-of-sight ($\cos i$ is uniformly distributed in the $[0, 1]$ interval), while w_{eff} is a small "effective" term used to reproduce the broadening effect of turbulence and non-circular motions on HI linewidths. The use of eqn. 2.4 is *only* justified if the HI disk is extended enough to sample the value of v_{rot} adopted by the model under consideration (e.g. v_{10kpc} for TG10 and $v_{rot,max}$ for Za09). We adopt the value $w_{eff} = 5\ \mathrm{km\,s^{-1}}$ for the broadening term[5], which is added linearly for galaxies with $v_{rot} > 50\ \mathrm{km\,s^{-1}}$ and in quadrature for lower velocity galaxies.

[5]The value of $w_{eff} = 5\ \mathrm{km\,s^{-1}}$ is derived empirically by Verheijen and Sancisi [2001], based on a sample of 22 galaxies with flat or decreasing outer rotation curves.

Figure 2.10 puts in evidence the marked departure of the theoretical distributions from the ALFALFA measurement at $w < 200\ \mathrm{km\,s^{-1}}$, which becomes more dramatic with decreasing width. According to the TG10 WF, the difference is approximately a factor of ~ 4 at $w = 100\ \mathrm{km\,s^{-1}}$, exhibiting an increasing trend. The Za09 WF[6], implies a difference of a factor of ~ 8 at the lowest width where the simulation is complete ($w \approx 50\ \mathrm{km\,s^{-1}}$), and displays a much steeper low-width slope than the ALFALFA measurement. An extrapolation of the Za09 WF to the ALFALFA width limit ($w = 20\ \mathrm{km\,s^{-1}}$), would result in a discrepancy of a factor of ~ 100.

2.5.2 Is CDM viable?

The ALFALFA measurement of the low-width end of the WF confirms the result obtained by the HIPASS survey [Zwaan et al., 2010], at lower sensitivity and velocity resolution. This fact excludes the possibility that the CDM overabundance problem is an artifact of the limited performance characteristics of past blind 21 cm surveys. The reason for the observed discrepancy can be therefore most likely attributed to one of the two following factors:

1. The inaccuracy of standard CDM simulations, presumably due to the inadequacy of the assumed DM model.
2. The improper comparison of simulated halos with observed galaxies. This could be due either to

[6]In order to account for the fact that the Za09 sample resides in an overdense volume (within a radius of $20\ h^{-1}$ Mpc from their simulated "Local Group"), we lower the normalization of their WF by a factor of 2, as suggested in their §4.3.

(a) the inadequate modeling of the baryonic counterparts hosted by DM halos, which leads to wrong predictions for the galactic rotation curve, or

(b) the incorrect interpretation of inclination-corrected HI linewidths as halo rotational velocities.

In what follows, we will consider these possibilities in more detail and argue about their prospects as solutions of the CDM overabundance problem.

Most large-scale simulations of cosmic structure conform to the standard ΛCDM cosmological model. In particular, they assume that all dark matter is cold (i.e. has negligible free-streaming length), non self-interacting and stable (i.e. non-decaying). These properties are appropriate for a universe where dark matter consists of stable weakly interacting massive particles (WIMPs). WIMPs are currently the favored DM particle candidate, and are expected to have masses in the GeV-TeV range and weak scale self-interaction cross-sections, justifying the DM attributes most commonly assumed in cosmological N-body simulations.

However, the picture changes considerably if DM is composed of relatively light ($\sim$ keV) particles, in which case it is referred to as warm dark matter (WDM). Structure on large scales would be the same as in a CDM universe, but on small scales halo formation would be heavily suppressed due to the non-negligible free-streaming length of the light WDM particle. Za09 have considered this alternative scenario, and carried out a second run of their very high-resolution simulation assuming a 1 keV WDM particle. They subsequently populate their halos with synthetic galaxies, employing the same modeling scheme as in their CDM run (§2.5.1). The result is shown by the red solid line in Fig-

ure 2.11, superposed on the ALFALFA WF (datapoints with errorbars and black solid line) and the result of their CDM run (blue solid line).

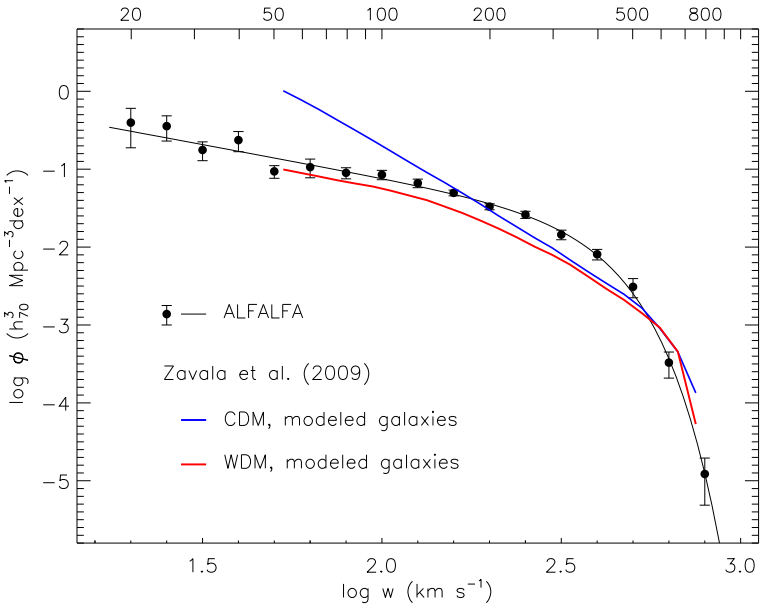

Figure 2.11 Data points with error bars and black solid line represent the measured ALFALFA WF (same as in Figure 2.4). The blue solid line represents the WF of a modeled galaxy population based on the high resolution CDM simulation of Zavala et al. [2009, Za09] (same as in Figure 2.10). The red solid line represents the WF corresponding to a second run of the Za09 simulation assuming a 1 keV WDM particle (both simulations employ the same scheme to populate halos with synthetic galaxies). The WDM WF displays a shallow low-width slope due to the suppressed formation of structure at small scales, and is in much better agreement with the ALFALFA measurement.

Strikingly, the synthetic WF in the WDM case exhibits a shallow slope at the low-width end, in good agreement with the slope measured by ALFALFA. Such a shallow slope results from the suppressed production of low-mass halos in a WDM universe, which directly translates into a lower abundance of low-width visible galaxies. WDM could therefore provide a simple and elegant solution of the overabundance problem.

Despite its appeal in this specific context, the general prospects of WDM also depend on its overall viability as the dominant constituent of non-baryonic

matter in the universe. A number of theoretical microscopic models for WDM have been proposed, most commonly involving sterile neutrinos [Dodelson and Widrow, 1994, Fuller et al., 2003, Asaka et al., 2005, Kusenko, 2009]. Constraints on the particle's mass can be placed by astrophysical and cosmological considerations. In particular, Lyα forest data place lower limits on the neutrino mass (a lighter particle generally results in suppression of power at larger scales), while X-ray observations can place upper mass limits (radiative decay into X-ray photons generally becomes more efficient at higher masses). The limits on the neutrino mass imposed by these observational constraints depend on the assumed neutrino production mechanism. Abazajian and Koushiappas [2006] find that non-resonantly produced neutrinos are ruled out, using a compilation of Lyα forest and X-ray data (see references therein). Boyarsky et al. [2009] have considered sterile neutrino production in the context of the νMSM (Minimal Standard Model + 3 sterile neutrinos) and argue that sterile neutrinos with $m_{sn} > 2$ keV are viable.

The second class of potential solutions attribute the disagreement between theory and observation to the process used to translate the output of simulations into actual galaxies. In particular, a number of important effects need to be taken into account (identified as items 1-4 in Section 2.5) to ensure a successful comparison of simulated halos with observed galactic samples.

Both theoretical works presented in §2.5.1 address issues 1 and 2. For example, Za09 set an explicit limit on the mass of halos hosting individual galaxies at $M_{vir} = 10^{13}\ h^{-1}\ M_{\odot}$. The influence of baryons on the shape of galactic rotation curves is also taken into account by both works, albeit using slightly different prescriptions and definitions of galaxy rotational velocity. Despite the use of nu-

merous simplifying assumptions by TG10 (e.g. all baryons within 10kpc) and Za09 (e.g. fixed disk-to-virial mass ratio for all galaxies) their theoretical WFs are in fair agreement with the ALFALFA measurement at intermediate widths ($200\ \mathrm{km\,s^{-1}} < w < 500\ \mathrm{km\,s^{-1}}$).

The last two issues are related specifically to the atomic hydrogen content of galaxies, which is not modeled by either TG10 or Za09. Specifically, issue 3 concerns the detectability of a galaxy in a 21cm survey. In principle, there exists the possibility that most of the low-mass halos predicted by CDM cosmology correspond to HI-devoid, dwarf spheroidal galaxies. In reality, a solution involving a multitude of isolated early-type dwarf systems seems rather unlikely. Direct observations [Garnett, 2002, Swaters and Balcells, 2002, Noordermeer et al., 2005], as well as other empirical arguments, suggest that the HI-to-stellar mass ratio grows with decreasing mass for galaxies in the field. HI surveys should thus have an advantage, rather than a disadvantage, at detecting the baryonic counterparts hosted by low-mass DM halos. In addition, optical surveys suggest that isolated early-type dwarfs in medium/low density environments are relatively rare [Karachentsev et al., 2004]. A second issue relates to the fact that satellite galaxies may be underrepresented in an HI-selected sample, since they are generally redder (and have presumably lower gas fractions) than central galaxies of the same luminosity [e.g. Font et al., 2008]. This bias could result in a $\lesssim 30\%$ underestimate of the abundance of low-width galaxies by ALFALFA [e.g Yang et al., 2008, Klypin et al., 2010], not nearly enough to explain the observed discrepancies.

Issue 4 regards the size and detailed spatial distribution of the atomic hydrogen component in galaxies, which determines the way in which its rotation

curve is converted into an HI velocity width. In particular, w_{HI} is a fair tracer of the maximum rotational velocity, only if the HI disk is extended enough to reach the flat (or decreasing) part of the galactic rotation curve. The use of Eqn. 2.4 in the derivation of the theoretical WFs implicitly assumes this situation to be true; observationally however, this is often times not the case. For example, the Catinella et al. [2006] set of template rotation curves, shows that lower rotational velocity galaxies tend to have steeper outer rotation curves (see their Figure 1 & 4). The dwarf galaxy samples of Spekkens et al. [2005] and Swaters et al. [2009], suggest that the effect becomes more dramatic at the lowest velocities (see Figure 3 & Figure 4 in the respective references).

This systematic trend for lower velocity galaxies to host less extended HI disks can be understood in terms of the expected baryon depletion of low-mass halos. Results from N-body + hydrodynamics simulations [e.g. Hoeft et al., 2006, Ricotti, 2010] indicate that halos with mass below some critical value lose a significant fraction of their cosmic share of baryonic matter, due to environmental and internal feedback processes. In particular, UV heating of the intergalactic medium (IGM) after reionization is believed to lead to substantial gas removal from low-mass halos ($v_{rot} \lesssim 20$ - $30\,\mathrm{km\,s^{-1}}$, corresponding to $M_{vir} \lesssim 10^9$ - $10^{9.5}\ h^{-1}\ M_{\odot}$). Internal feedback processes such as supernova winds may also be important, but their efficacy is strongly model dependent.

The above considerations could lead to a solution of the overabundance problem that would not require a modification of the extremely successful ΛCDM paradigm. In simple terms, the overabundance problem would be the result of the inability of HI to trace the maximum halo rotational velocity of low-mass systems, which leads to a severe underestimate of their true mass.

The same argument has been identified as a possible solution of the "mini-void size" problem by Tikhonov and Klypin [2009], while a similar effect has been proposed by Peñarrubia et al. [2008] as a solution to the "missing satellites" problem.

2.6 The v_{rot} - v_{halo} relation in a CDM universe

Assuming the CDM model to be correct, we can statistically infer the v_{rot} - v_{halo} relationship needed to reproduce the observational galaxy VF. This can be done by abundance matching, a statistical procedure which assumes the existence of a one-to-one relationship between galaxy and halo circular velocities, $v_{rot} = f(v_{halo})$. It follows that the space density of halos with circular velocities larger than a given value, V, should be equal to the space density of galaxies with rotational velocities larger than the value dictated by the relationship, $n(v_{halo} > V) = n(v_{rot} > f(V))$.

Obtaining an observational velocity distribution from the ALFALFA measurement is not straightforward. Firstly, the ALFALFA measurement regards galaxy velocity widths (uncorrected for inclination) and not intrinsic rotational velocities; secondly, the ALFALFA survey is biased against HI-poor massive ellipticals that dominate the counts at high velocities.

We address the first issue by searching for the velocity distribution that best reproduces the ALFALFA WF, upon projection using Eqn. 2.4. We assume that the distribution follows a modified Schechter function of the form

$$\phi(v) = \frac{dn}{d\log v} = \ln(10)\,\phi_* \left(\frac{v}{v_*}\right)^{\alpha} e^{-(\frac{v}{v_*})^{\beta}} \quad , \tag{2.5}$$

and that it corresponds to the VF of HI-rich, late-type galaxies. The set of parameters that provide the best match is identified visually, and corresponds to the values $\phi_* = 1.2 \cdot 10^{-2}\ h_{70}^3\ \mathrm{Mpc}^{-3}$, $\log v_* = 2.32$, $\alpha = -0.81$ and $\beta = 3.1$ (thick red dash-dotted line in Figure 2.12). In order to address the second issue (i.e. obtain a VF valid for all morphological types), we use the results of Chae [2010], who studied the velocity dispersion function (VDF) of early-type galaxies in the SDSS and 2dFGRS surveys. Velocity dispersions can be transformed into rotational velocities by assuming an isothermal mass profile, in which case $v_{rot} = \sqrt{2}\sigma$. We adopt the average of the 2dFGRS and SDSS velocity distributions as a representative VF for early-type galaxies, which we plot as the green dotted line in Figure 2.12.

We interpolate the two distributions using a single modified Schechter function with parameters $\phi_* = 8.7 \cdot 10^{-3}\ h_{70}^3\ \mathrm{Mpc}^{-3}$, $\log v_* = 2.49$, $\alpha = -0.81$ and $\beta = 3.35$. The interpolated distribution (blue solid line in Figure 2.12 & Figure 2.13) represents a composite galactic VF valid for all morphological types. Even though we do not formally measure errors for the derived distribution, we list below some important sources of uncertainty. Firstly, the statistical uncertainty on the parameters of the late-type VF should be at least on the order of the errors reported in §2.3.1; the parameters of the composite VF should be expected to carry larger errors, since the determination of the interpolating distribution is subjective to some extent. More importantly though, there are a number of sources of uncertainty related to galactic physics. For example, the assumption of isothermality of early-type galaxies is not expected to hold in

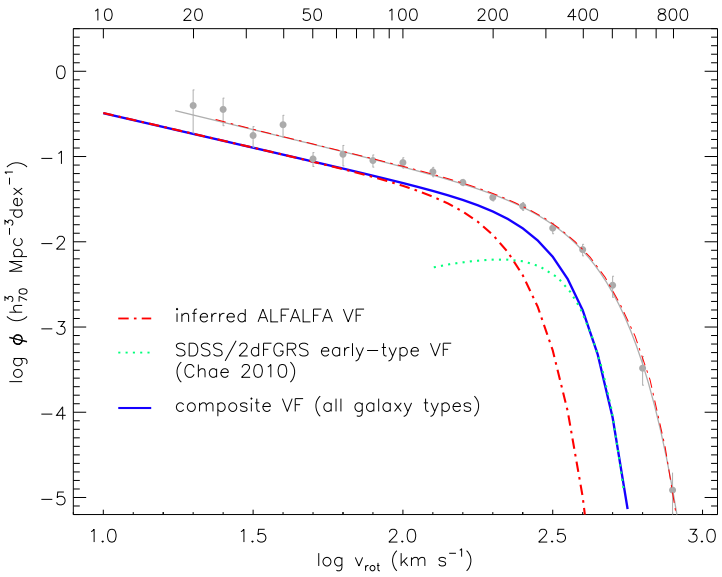

Figure 2.12 The thick red dash-dotted line represents the velocity function of late-type galaxies (assumed to follow a modified Schechter distribution) that best reproduces the measured ALFALFA WF (light gray data points and solid line) upon projection (thin red dash-dotted line). The green dotted line represents the velocity function of early-type galaxies determined by Chae [2010] using SDSS and 2dFGRS data. The blue solid line is a modified Schechter interpolation of the two VFs which represents a velocity function valid for all morphological types. The modified Schechter parameters for the interpolated distribution are $\phi_* = 8.7 \cdot 10^{-3}\ h_{70}^3\ \mathrm{Mpc}^{-3}$, $\log v_* = 2.49$, $\alpha = -0.81$ and $\beta = 3.35$ (see Section 2.6 for more details).

detail [e.g Dutton et al., 2010], which would affect the high-velocity end of the composite VF. Moreover, the low velocity slope, α, depends partly on the value of w_{eff} employed in Eqn. 2.4; the value adopted here ($w_{eff} = 5\ \mathrm{km\,s^{-1}}$) has been empirically determined from a sample of relatively massive spirals [Verheijen and Sancisi, 2001], and does not have to be the same for galaxies populating the low-velocity end of the VF. Also, as mentioned in §2.5.2, the inferred VF may be underestimating the true abundance of low-width galaxies by $\lesssim 30$ %, since ALFALFA is likely to miss some fraction of the satellites of massive spiral galaxies.

Next, we obtain the theoretical CDM VF from the Bolshoi simulation[7] [Klypin et al., 2010]. In particular, we use the distribution of maximum halo rotational velocity, v_{halo}, of all simulated halos (including subhalos) at the present epoch, which is shown as the black solid line in Figure 2.13. Note that the simulation is run for the total matter density of the universe ($\Omega_m = \Omega_{DM} + \Omega_{bar} = 0.27$), but both DM and baryons are treated as dissipationless components. Also note that the simulation is complete only down to $v_{halo} = 50\ \mathrm{km\,s^{-1}}$, and a power-law extrapolation is used at lower velocities [which is however expected to hold, see for example §4.2 in Zavala et al., 2009].

The red thick line in Figure 2.14 represents the v_{rot} - v_{halo} relation obtained by matching the CDM and galactic velocity distributions (values listed in Table 2.1). We have assumed that halos with $v_{halo} > 360\ \mathrm{km\,s^{-1}}$ ($M_{vir} \gtrsim 10^{13}\ h^{-1}\ M_{\odot}$) do not host individual galaxies but rather groups of galaxies, and are hence excluded from the matching process. The cyan shaded region corresponds to different values for this mass cutoff, ranging from $v_{halo,max} = 290\ \mathrm{km\,s^{-1}}$ ($M_{vir} \approx$

[7]The Bolshoi simulation is run for the set of cosmological parameters $h = 0.70$, $\Omega_m = 0.27$, $\sigma_8 = 0.82$, $n = 0.95$.

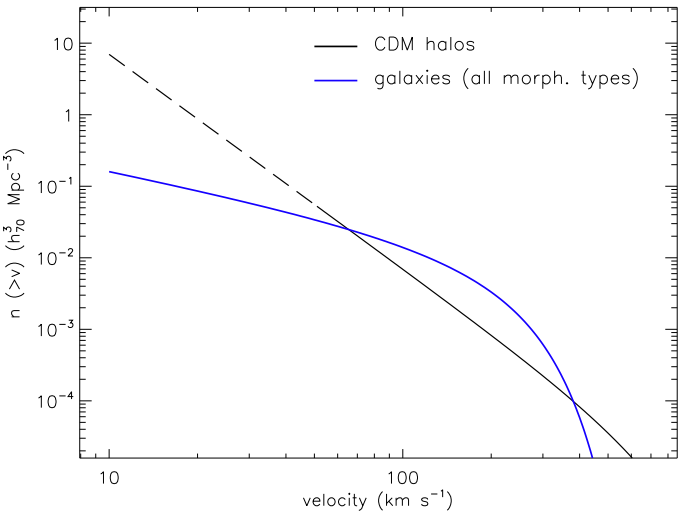

Figure 2.13 The velocity function of halos (black line) and galaxies (blue line), expressed as a cumulative distribution. The former distribution corresponds to the number density of halos (including subhalos) in the Bolshoi CDM simulation, as a function of their maximum rotational velocity at the present epoch (v_{halo}). Note that the Bolshoi simulation is complete only down to $v_{halo} = 50\ \mathrm{km\,s^{-1}}$, but a power-law extrapolation to lower velocities (black dashed line) is expected to hold. The latter distribution represents the VF of all galaxy types, as a function of their observed rotational velocity (same as blue line in Figure 2.12, see discussion is Section 2.6).

$5 \cdot 10^{12}\ h^{-1}\, M_{\odot}$, upper boundary) to $v_{halo,max} = 440\ \mathrm{km\,s^{-1}}$ ($M_{vir} \approx 2 \cdot 10^{13}\ h^{-1}\, M_{\odot}$, lower boundary). The uncertainty in the value of $v_{halo,max}$ mentioned above is the only source of error considered explicitly here. There are, however, additional uncertainties involved in the determination of the presented relationship. For example, no scatter in the $v_{rot} = f(v_{halo})$ relation was considered in the abundance matching process. Also, no corrections to v_{rot} for pressure support have been made in this work, even though gas thermal velocities in low mass galaxies can be comparable with their rotational velocities.

Figure 2.14 shows that v_{rot} follows an approximately linear relationship with v_{halo} only for intermediate-mass halos ($120\ \mathrm{km\,s^{-1}} \lesssim v_{halo} \lesssim 170\ \mathrm{km\,s^{-1}}$). In this range, $v_{rot} \approx 1.5\, v_{halo}$, in fair agreement with the values estimated for the MW and M31 from dynamical models [Klypin et al., 2002, *diamonds]* and from the kinematics of MW high velocity stars [Smith et al., 2007, *triangle]* and blue horizontal branch stars [Xue et al., 2008, *box]*. However, the v_{rot}/v_{halo} ratios obtained here are significantly larger than the average values inferred by Dutton et al. [2010] from a compilation of weak lensing and satellite kinematics datasets. Note though that their results are expressed in terms of a v_{opt} - v_{200} relation, where v_{opt} is defined as the measured rotational velocity at 2.2 *I*-band disk scalelengths for late-type galaxies and 1.65σ for early-type galaxies, and v_{200} refers to the virial velocity of the halo at an enclosed overdensity of 200 times the critical density. In order to display their results in Figure 2.14, (red & blue hatched regions) we have transformed virial velocities into halo maximum rotational velocities assuming average halo concentrations [Macciò et al., 2008].

The most important result of Figure 2.14 concerns the low halo velocity regime. In particular the relationship steepens continuously as we move to

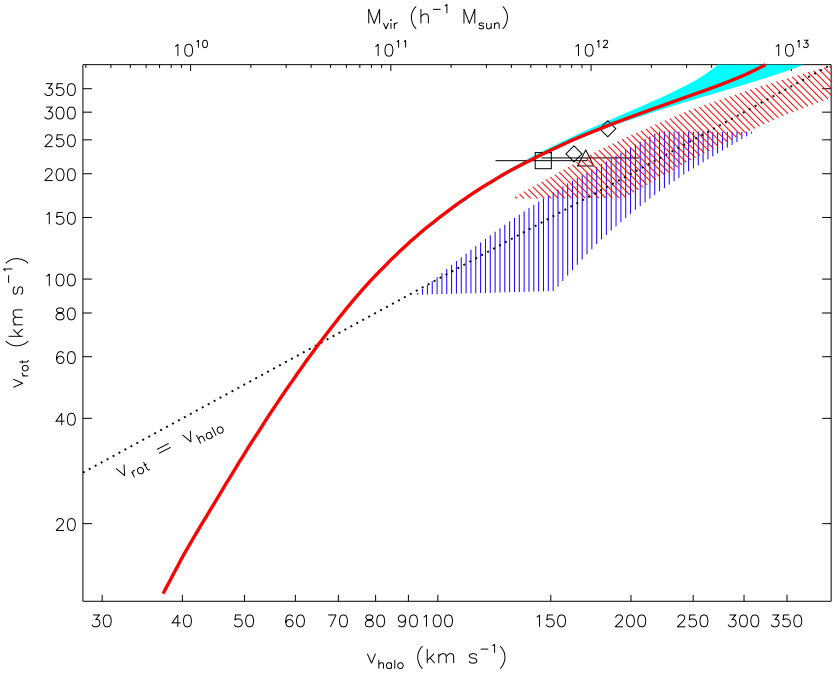

Figure 2.14 The red solid line corresponds to the relation between the rotational velocity of galaxies measured observationally (v_{rot}) and the maximum rotational velocity of the corresponding CDM halo (v_{halo}). The relation was obtained by the abundance matching of the velocity distribution of halos in the Bolshoi CDM simulation with the velocity distribution of galaxies inferred from ALFALFA and SDSS/2dFGRS data (see Figure 2.12 & 2.13). We have assumed that halos with $v_{halo} > 360\ \mathrm{km\,s^{-1}}$ ($M_{vir} \gtrsim 10^{13}\ h^{-1}\,M_{\odot}$) do not host individual galaxies, but rather groups of galaxies. The cyan shaded area corresponds to different mass cutoffs, ranging from $v_{halo,max} = 290\ \mathrm{km\,s^{-1}}$ (*upper boundary*) to $v_{halo,max} = 440\ \mathrm{km\,s^{-1}}$ (*lower boundary*). The blue and red hatched areas correspond to the 2σ error regions for late- and early-type galaxies respectively, according to Dutton et al. [2010]. Their measurement was based on a compilation of weak lensing and satellite kinematics measurements of galaxy dynamical masses (see Section 2.6 for more details). The symbols correspond to the values estimated for the MW and M31 based on dynamical models [Klypin et al., 2002, diamonds], and for the MW based on the kinematics of high velocity stars [Smith et al., 2007, triangle with 2σ errorbars] and blue horizontal branch stars [Xue et al., 2008, box with 2σ errorbars].

Table 2.1. The v_{rot} - v_{halo} relation in a CDM universe

v_{halo} [$\mathrm{km\,s^{-1}}$]	v_{rot} [$\mathrm{km\,s^{-1}}$] ($v_{halo,max} = 360\,\mathrm{km\,s^{-1}}$)	v_{rot} [$\mathrm{km\,s^{-1}}$] ($v_{halo,max} = 290\,\mathrm{km\,s^{-1}}$)	v_{rot} [$\mathrm{km\,s^{-1}}$] ($v_{halo,max} = 440\,\mathrm{km\,s^{-1}}$)
40	16	16	16
45	23	23	23
50	32	32	32
55	42	42	42
60	53	53	53
70	77	78	77
80	102	103	102
90	127	127	125
100	149	150	147
120	188	190	185
140	219	223	218
160	247	252	244
180	270	278	267
200	291	303	286
220	310	327	303
240	328	353	318
260	345	383	333
300	383	⋯	360
340	431	⋯	387
380	⋯	⋯	416
420	⋯	⋯	449

Note. — v_{halo} corresponds to the maximum circular velocity of a halo (including subhalos) in the Bolshoi simulation [Klypin et al., 2010], at $z = 0$. Note that the simulation is run for the total matter density of the universe ($\Omega_m = \Omega_{DM} + \Omega_{bar} = 0.27$). We have excluded halos with velocities larger than $v_{halo,max}$ from the abundance matching procedure, since these are assumed to host groups of galaxies. v_{rot} refers to the rotational velocity of galaxies inferred observationally.

lower halo velocities, assuming a power-law behavior of the form $v_{rot} \propto v_{halo}^3$ at $v_{halo} \lesssim 50\ \mathrm{km\,s^{-1}}$. As a result, the true mass of low-velocity halos is systematically underestimated when measured by the inclination-corrected HI linewidth of the hosted galaxy; the underestimate can reach a factor of ~ 2.5 for $v_{halo} \approx 40\ \mathrm{km\,s^{-1}}$. Testing the low-velocity end of the v_{rot} - v_{halo} relation would require a sample of low-mass galaxies with directly measured dynamical masses, e.g. through weak lensing or satellite kinematics. However, some indirect observational support could come from a rich sample of HI interferometric maps of dwarf galaxies: a gradual transition from mostly flat to mostly rising rotation curves at $v_{rot} \approx 110\ \mathrm{km\,s^{-1}}$, would be required to explain the steepening of the relation at low velocities. Ultimately, observational verification of the presented relationship at low velocities would provide a check of the validity of the CDM model.

2.7 Conclusions

We have measured the velocity width function (WF) of HI-bearing galaxies, based on a sample of $10\,744$ extragalactic sources detected in ∼40% of the final ALFALFA survey area. The ALFALFA measurement extends to widths (uncorrected for inclination) as low as $w = 20\ \mathrm{km\,s^{-1}}$, and results in a robust measurement of the low-width logarithmic slope of $\alpha = -0.85 \pm 0.19$ (1σ statistical error including the effect of cosmic variance). This result suggests a significant incompatibility of the observational distribution with the much steeper distribution expected in a CDM universe.

We compare the ALFALFA result with the WFs of two modeled galaxy popu-

lations, one populating the Bolshoi CDM simulation halos [Trujillo-Gomez et al., 2010] and the other populating the halos of the very-high-resolution CDM simulation of Zavala et al. [2009]. Indeed, the simulated WFs start diverging from the ALFALFA measurement at widths $w \lesssim 200\ \mathrm{km\,s^{-1}}$. The difference in abundance is a factor of ~ 8 at $w = 50\ \mathrm{km\,s^{-1}}$ (which corresponds to the resolution limit of the Za09 simulation), and implies a difference of a factor of ~ 100 when extrapolated to the ALFALFA low-width limit ($w = 20\ \mathrm{km\,s^{-1}}$). This discrepancy is closely related to a number of other observational challenges to CDM (e.g. "missing satellites problem", "mini-void size problem", etc.), which we collectively refer to as the *CDM overabundance problem*.

We further identify the two most promising solutions to the problem: the first involves the suppression of low-mass halo formation, which is best accomplished by assuming a ∼keV WDM particle; the second solution does not require a modification of the extremely successful CDM model, and relies on the fact that HI disks in dwarf galaxies are frequently not extended enough to probe the full amplitude of the galactic rotation curve. The latter solution, supported by currently limited observational evidence, implies that galaxy rotational velocities derived from inclination-corrected HI linewidths (v_{rot}) systematically underestimate the maximum rotational velocity of their host DM halo (v_{halo}), below $v_{rot} \approx 110\ \mathrm{km\,s^{-1}}$.

We furthermore use an abundance matching procedure to statistically infer the v_{rot} - v_{halo} relationship needed to reconciliate the CDM and galactic velocity distributions. We find that for MW-sized galaxies $v_{rot} \approx 1.5\ v_{halo}$, while at low velocities v_{rot} underestimates significantly the true maximum rotational velocity of the host halo.

Determining the correct solution to the CDM overabundance problem rests both on the general prospects of WDM as a viable dark matter model, as well as on observational verification of the v_{rot} - v_{halo} relationship predicted for CDM. The latter goal could be best accomplished through a rich sample of low-mass galaxies with directly measured dynamical masses.

CHAPTER 3

A DIRECT MEASUREMENT OF THE BARYONIC MASS FUNCTION OF GALAXIES & IMPLICATIONS FOR THE GALACTIC BARYON FRACTION

3.1 Introduction

It is by now well established that baryonic matter represents only about $1/6$ of the total matter density of the universe [e.g. Komatsu et al., 2011], while the majority is in the form of non-baryonic dark matter (DM). Since galaxies form through the accretion of baryonic material onto dynamically dominant DM structures (halos), it would be reasonable to assume that the baryon mass fraction of present day galaxies approximately equals the cosmic value ($f_b = \Omega_b/\Omega_m \approx 0.16$). Despite this expectation, observations point to the fact that galaxies are not able to retain their cosmic "fair share" of baryons, and that the resulting baryon deficit depends strongly on the mass of their host halo.

The first line of evidence is provided by observational estimates of the abundance of galaxies as a function of their total stellar mass, a distribution referred to as the galactic stellar mass function (SMF). Thanks to the advent of wide area optical surveys with multiband photometric and spectroscopic information, such as the Two degree Field Galaxy Redshift Survey (2dFGRS) and the Sloan Digital Sky Survey (SDSS), the SMF has been measured over the mass range $M_* \approx 10^7 - 10^{12}\ M_\odot$, using statistical samples of tens of thousands of

*This chapter is an adapted version of the published article Papastergis et al. (2012).

galaxies and a variety of stellar mass estimation techniques [Cole et al., 2001, Bell et al., 2003, Panter et al., 2007, Baldry et al., 2008, Li and White, 2009, Yang et al., 2009, Baldry et al., 2012, to name a few]. The SMF displays an exponential cutoff at masses $M_* \gtrsim 10^{11} M_\odot$ and an approximate power-law behavior at low masses ($dn \propto M_*^{-\alpha}\, dM_*$), with a "shallow" exponent of $\alpha \approx -1.3$. On the other hand, the halo mass function (HMF) predicted in the lambda cold dark matter (ΛCDM) model, follows a much "steeper" power-law ($\alpha \approx -1.9$) over the mass range of interest. This observation alone excludes the possibility that the stellar mass of a galaxy is simply a fixed fraction of the host halo mass.

One can furthermore statistically derive an average relation between the stellar mass of a galaxy (M_*) and the mass of its host halo (M_h), through the technique of abundance matching (see §3.5.1 for details). M_* - M_h relations based on abundance matching [e.g. Guo et al., 2010, Moster et al., 2010, Behroozi et al., 2010, Leauthaud et al., 2012] have shown that the "stellar conversion efficiency", $\eta_* = (M_*/M_h) \,/\, f_b$, never exceeds 25 - 30%. Furthermore, η_* peaks for Milky Way-sized galaxies ($M_h \approx 10^{12} M_\odot$), and declines rapidly on either side of the peak [e.g. Figure 2 in Guo et al., 2010].

The second line of evidence comes from direct halo mass measurements, obtained through weak lensing or kinematics studies [e.g. Dutton et al., 2010, Reyes et al., 2012]. For example, Reyes et al. [2012] used stacked weak lensing measurements to estimate the average host halo mass of galaxies in different stellar mass bins, and found that η_* never exceeds $\approx 30\%$. Direct halo mass measurements can circumvent a number of assumptions inherent in the application of abundance matching, but such techniques can presently only be applied to a restricted range of stellar mass ($M_* \approx 10^9 - 10^{11}\ M_\odot$), and are affected by their

own set of systematics.

Stellar mass is not always the dominant baryonic component in a galaxy. In fact, the HI–to–stellar mass ratio ("HI fraction"; $f_{HI} = M_{HI}/M_*$) tends to increase with decreasing stellar mass, and HI often dominates the baryonic content of low-mass galaxies. The transition from stellar-mass–dominated to HI–dominated systems takes place at $M_* \approx 10^{10} M_\odot$ for HI-selected samples [e.g. Huang et al., 2012a, see also Fig. 3.19 in this work], or at $M_* \lesssim 10^{9.5}\ M_\odot$ for optically-selected samples [e.g. Catinella et al., 2010]. As a result, it is presently not clear what is the behavior of the "baryon retention fraction" $\eta_b = (M_b/M_h)\ /\ f_b$ in low-mass galaxies, when both stars and cold gas are taken into account. In particular it is not well understood whether the very low average value of η_* inferred for low-mass halos is a result of poor retention of baryonic material, of the low efficiency of gas-to-stars conversion, or of a combination of both.

For example, Baldry et al. [2008] argue that the increasing gas fraction in low-mass galaxies should approximately offset the decreasing stellar-to-halo mass ratio, and result in a roughly constant $\eta_b \approx 10\%$. This conclusion was based on an indirect estimate of the cold gas content of galaxies, based on the average $f_{HI} - M_*$ relation observed in a set of samples in the literature. An early work by Salucci and Persic [1999], based on the same indirect method, also reached a qualitatively similar conclusion. Evoli et al. [2011] found an approximately constant η_b at the low-mass end using a different indirect method, which involves the comparison of the stellar and HI mass distributions of two different galaxy samples. These results would imply that low-mass galaxies are relatively efficient at retaining baryonic mass, but very inefficient in converting

their gas into stars. This conclusion, however, would require a "steep" HI mass function (HIMF) in the local universe, in contrast to what is measured [Zwaan et al., 2005, Martin et al., 2010]. Moreover, the recent work of Rodríguez-Puebla et al. [2011], also based on using the average $f_{HI} - M_*$ relations for blue and red galaxies separately, found no signs for a flat η_b at low masses.

In this chapter we *directly* measure the abundance of galaxies as a function of their "baryonic mass" (throughout this chapter the term baryonic refers to the combined stellar and atomic gas components of galaxies, and baryonic mass is calculated as $M_b = M_* + 1.4\,M_{HI}$, where the 1.4 factor accounts for the presence of helium). We use optical data from the seventh data release of the SDSS (SDSS DR7) to estimate stellar masses, and HI-line flux measurements from the Arecibo Legacy Fast ALFA[1] (ALFALFA) survey to measure atomic gas masses. The resulting distribution, referred to hereafter as the baryon mass function (BMF) of galaxies, can be used in abundance matching to derive a robust η_b - M_h relation. In order to investigate sample selection effects, we employ both an HI-selected and an optically-selected sample drawn from the *same* volume to derive the mass distributions for the stellar, atomic hydrogen and baryonic components.

The chapter is organized as follows: in section 3.2, we introduce the datasets used to measure the stellar, HI and baryon mass distributions. We describe the methodology used to measure atomic hydrogen masses and we estimate stellar masses for our galaxy samples. In section 3.3, we present our measurements of the SMF, HIMF & BMF from both the HI-selected and the optically-selected samples, and compare them against one another as well as against other pub-

[1]The Arecibo L-band Feed Array (ALFA) is a 7-feed receiver operating in the L-band (≈ 1420 MHz), installed at the Arecibo Observatory.

lished results. In section 3.4 we consider the impact of possible systematics on our measurements, such as stellar mass estimation method, distance uncertainties and the exclusion of some baryonic components (e.g. molecular gas) in the calculation of the BMF. In section 3.5, we present the η_* - M_h and η_b - M_h relation in a ΛCDM universe. In section 3.6, we discuss the implications of the result and summarize our main conclusions. We note again that throughout this chapter we use a Hubble constant of $H_0 = 70\,h_{70}\ \mathrm{km\,s^{-1}Mpc^{-1}}$.

3.2 Datasets & derived quantities

3.2.1 HI-selected sample

We select galaxies from the current data release of the ALFALFA survey, which covers 40% of the planned final survey area ["α.40" catalog; Haynes et al., 2011]. We restrict ourselves to two rectangular areas of the "spring" coverage of α.40 ($07^h45^m < RA < 16^h30^m$, $4^\circ < Dec < 16^\circ$ & $24^\circ < Dec < 28^\circ$), which encompass the Virgo cluster as well as the supergalactic plane at low velocities. We restrict ourselves to galaxies with $v_{CMB} < 15\,000\ \mathrm{km\,s^{-1}}$ ($z < 0.05$), in order to avoid the strong radio frequency interference (RFI) present at frequencies that correspond to $v_\odot \gtrsim 15\,000\ \mathrm{km\,s^{-1}}$. We discard the nearest extragalactic sources with $D < 10$ Mpc, because they can carry extreme fractional uncertainties on their distances (see §3.2.3 for details on the distance assignment method). We furthermore select only HI sources designated as "Code 1" in α.40, i.e. extragalactic sources detected at high significance ($S/N_{HI} > 6.5$). In addition, we exclude sources with integrated fluxes below the 50% completeness limit of the

ALFALFA survey [see Haynes et al., 2011, Section 6 for the derivation of the ALFALFA completeness limits]. The above requirements are satisfied by $7\,618$ galaxies.

We remove from our sample 204 α.40 sources which are not crossmatched with an optical source in SDSS, as well as 208 additional sources which have been flagged as having problematic SDSS photometry (crossmatch code "P" in α.40, for details see Section 4 in Haynes et al., 2011). This quality cut on the SDSS photometry introduces some bias against faint, low surface brightness galaxies of irregular morphology; such sources are often "shredded" (i.e. assigned multiple photometric objects) by the SDSS magnitude extraction process, and are usually assigned a "P" ("photometry suspect") crossmatch code in α.40. Lastly, 11 additional objects were discarded, in cases where the stellar mass computation method described in §3.2.3 failed to produce physically plausible results.

Our final sample thus consists of $7\,195$ extragalactic objects, detected over $\approx 2\,000$ deg^2 of high Galactic latitude sky and out to $D \approx 214$ Mpc. The upper panel of Figure 3.1 displays the spatial distribution of our HI-selected galaxies, and puts in evidence the complex large scale structure in the survey volume. Note that all objects in our HI-selected sample have 21cm redshifts [2] and line fluxes as well as multi-band optical photometry, and hence estimates of both their stellar and atomic hydrogen masses.

[2] Of the $7\,195$ galaxies in the HI-selected sample, $1\,333$ are not in the SDSS DR7 spectroscopic database and thus lack SDSS optical redshifts.

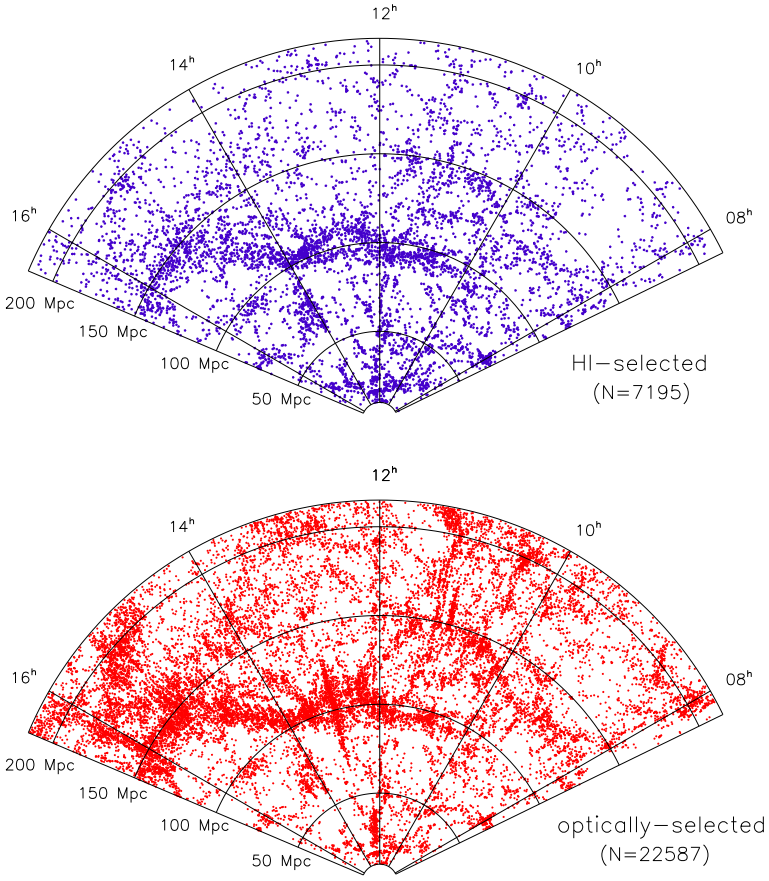

Figure 3.1 Spatial distribution of the 7 195 HI-selected galaxies (upper panel) and 22 587 optically-selected galaxies (lower panel), drawn from the same volume. The galaxy stellar mass function (SMF), HI mass function (HIMF) and baryonic mass function (BMF) are computed separately for the two samples, in order to assess the impact of sample selection on the derived distributions.

3.2.2 Optically-selected sample

We draw an optically-selected sample from the SDSS DR7 [Abazajian et al., 2009] spectroscopic database, in the same volume used to define our HI-selected sample. Specifically, we select galaxies that lie within the same sky area ($07^h45^m < RA < 16^h30^m$, $4^\circ < Dec < 16^\circ$ & $24^\circ < Dec < 28^\circ$) and satisfy the same velocity and distance restrictions ($v_{CMB} < 15\,000\,\mathrm{km\,s^{-1}}$, $D > 10$ Mpc). CMB velocities for our optically-selected galaxies are calculated based on their SDSS spectroscopic redshifts (z_{SDSS}). We restrict ourselves to objects spectroscopically classified as galaxies in SDSS (`specClass = 2`) that also have an apparent Petrosian magnitude brighter than 17.5 in the r-band ($r_{petro} < 17.5$). This initial cut results in $22\,707$ galaxies. Due to their large number, it is not practical to inspect all galaxies individually for the quality of their SDSS photometry/spectroscopy. As a result, we expect a fraction of our sources to have issues with their SDSS photometry, most often related to "shredding" (i.e. assignment of multiple photometric objects to a single galaxy). This issue affects mostly extended sources with structure in their light distribution, such as low surface brightness (LSB) galaxies with irregular morphology. In such cases, the SDSS magnitude will underestimate the true flux of the galaxy, which in turn will result in an underestimate of its stellar mass. A second issue related to shredding, is that bright star forming knots in the disks of nearby spiral galaxies can sometimes be cataloged as separate spectroscopic objects, and hence interpreted as low-mass satellites of the main spiral. We find that applying a color cut on our sample, $(i-z)_{model} > -0.25$, removes a fair fraction of these unwanted cases. On the other hand, cuts based on the quality of the SDSS spectrum (such as cuts on `zconf`, `zstatus` or `zwarning`) are ineffective, since they exclude mostly legitimate faint or LSB dwarf galaxies with noisy spectra. Lastly, we ex-

clude objects for which the stellar mass computation described in §3.2.3 failed to produce physically plausible results.

Our final optically-selected sample consists of $22\,587$ galaxies, occupying the same volume as our HI-selected sample. We crossmatch the optical sample with the full α.40 catalog (including Code 1 & 2 sources), and find $7\,551$ HI source counterparts. The crossmatch rate is thus approximately $1/3$, as reported in Haynes et al. [2011]. Note that ALFALFA non-detected galaxies are not necessarily HI-poor objects; due to the low emissivity of atomic hydrogen in the 21cm line, even moderately gas-rich galaxies in the outer portion of the survey volume can be missed by ALFALFA. This point is illustrated by the lower panel of Figure 3.1, which compares the spatial distribution of galaxies in the HI-selected and optically-selected samples. Note that all optically-selected galaxies have multiband optical photometry as well as optical redshifts, and hence an estimate of their stellar mass. However, only galaxies crossmatched with an α.40 source have a 21cm flux measurement, and hence a value for their atomic hydrogen mass.

3.2.3 Derived quantities

We calculate HI masses from the measured 21cm integrated flux reported in α.40. Assuming optically thin emission

$$M_{HI} = 2.356\,10^5\ S_{int}\ D^2\ , \tag{3.1}$$

where M_{HI} is the HI mass in units of the solar mass ($M_\odot$), S_{int} is the integrated

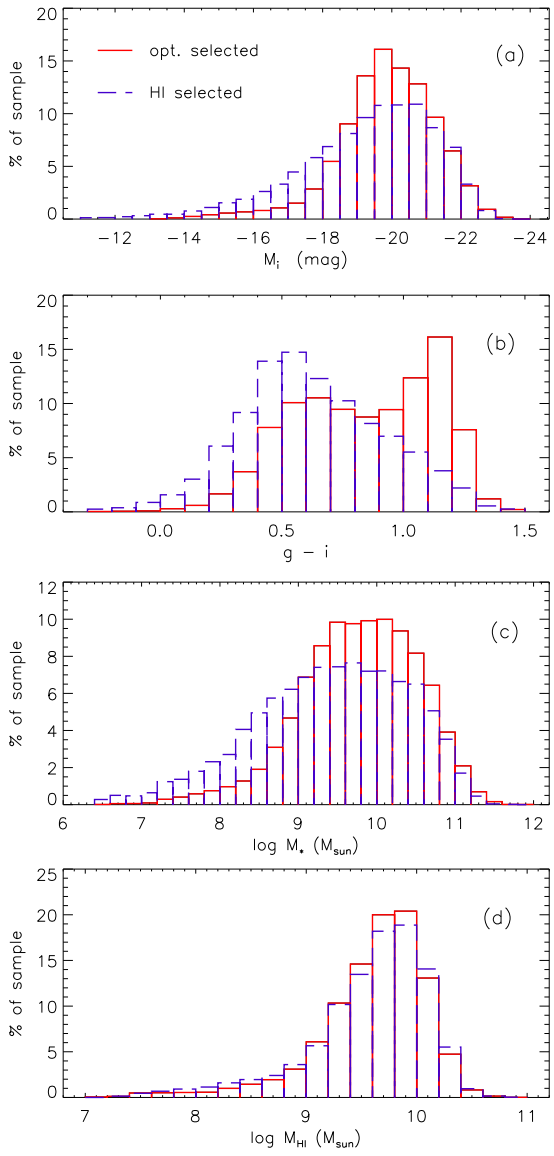

Figure 3.2 Histograms of i-band absolute magnitude (panel *a*), $g - i$ color (panel *b*), stellar mass (panel *c*) and HI mass (panel *d*), for the optically-selected (red solid line) and HI-selected (blue dashed line) samples. As evident in panel *b*, the HI-selected sample is strongly biased against the red galaxy population and as a result it is skewed towards lower luminosity and stellar mass systems. Conversely, the fractional contribution of bright and massive galaxies ($M_i \lesssim -19,\ M_* \gtrsim 10^9 M_\odot$) is larger for the optically-selected sample (panels *a* and *c*).

flux in Jy km s^{-1} and D is the distance in Mpc. Distances in this chapter are calculated according to the method used in α.40 [Haynes et al., 2011]: nearby galaxies ($v_{CMB} < 6\,000$ km s^{-1}) are assigned distances through the use of a peculiar velocity flow model developed by Masters [2005], while for more distant galaxies simple Hubble distances are used ($D = v_{CMB}/H_0$, with $H_0 = 70$ km s^{-1} Mpc^{-1}). Moreover, group and cluster membership information is taken into account when available, as well as primary distance measurements published in the literature. We would like to point out that most of the galaxies in our optically-selected sample are not included in α.40, and hence lack the systematic group assignments and primary distance information contained in the catalog. Nevertheless, optically-selected galaxies that lie within the sky area and redshift range of the Virgo cluster are placed collectively at the Virgo distance ($D = 16.5$ Mpc), in order to minimize the effects of peculiar motions on the inferred distances of galaxies in the region. We would also like to note that the distance assignment method can have a large impact on the determination of mass functions, especially at the low-mass end. We illustrate this issue in §3.4.2, where we consider the effect on the HIMF of using uniformly Hubble distances for all galaxies.

We compute stellar masses for our galaxies based on fitting all 5 SDSS photometric bands (u, g, r, i, z), with model spectral energy distributions (SEDs). The full details of the method can be found in Huang et al. [2012b], but here we summarize the main points: a library of model SEDs are generated, using the Bruzual and Charlot [2003] stellar population synthesis code and assuming a Chabrier [2003] stellar initial mass function (IMF). Models with an extensive range of internal extinction, metallicity and star formation histories are considered. In particular, star formation history templates include both an expo-

nentially declining component as well as random starburst episodes. The final physical properties (e.g. stellar mass, star formation rate, internal extinction etc.) are computed as the average of all model values, where each model is weighted according to its fit likelihood. In addition to mean values, "1σ" uncertainties of the physical properties can also be derived, as one quarter of the 2.5-97.5 percentile range of model values. The median 1σ uncertainty in $\log M_*$ is 0.086 dex, or about 22% (excluding uncertainties on the distance). It is important to note that stellar mass estimates of the same galaxy obtained with different methods can have systematic offsets of up to factors of a few. In §3.4.1 we address issues related to stellar mass estimation, and consider alternative methods for calculating stellar masses [Bell et al., 2003, Taylor et al., 2011].

Figure 3.2 compares the distributions of i-band absolute magnitude (M_i) and $g-i$ color (both corrected for Milky Way extinction), stellar mass (M_*) and HI mass (M_{HI}) for the HI-selected and optically-selected samples. The most notable difference is in the $g-i$ color distribution, where the HI-selected sample shows a strong bias against the red galaxy population [see also Huang et al., 2012a]. As a result, the optically-selected sample contains a larger proportion of high luminosity and stellar mass systems compared to the HI-selected sample. By contrast, the M_{HI} distribution of the two samples is very similar, but remember that only those optically-selected galaxies that are detected in ALFALFA are included in the histogram.

3.3 The baryonic mass function

3.3.1 Method

Stellar masses for all galaxies, and HI masses for all ALFALFA-detected galaxies are calculated as described in §3.2.3. For the $\approx 15\,000$ galaxies in the optically-selected sample that lack an ALFALFA detection, we assign a lower and an upper limit on their atomic hydrogen content ($M_{HI}^{min}, M_{HI}^{max}$). The lower limit is simply $M_{HI}^{min} = 0$, which corresponds to an HI-devoid galaxy. The upper limit is calculated by assuming that the HI flux of the non-detected galaxy lies just below the ALFALFA "detection limit", as defined by the 25% completeness limit of the α.40 catalog when both Code 1 & 2 sources are considered. More precisely

$$\log M_{HI}^{max} = 5.372 + \log S_{int}^{25\%\,lim} + 2\log D \ , \tag{3.2}$$

where D is the galaxy distance in Mpc determined by its SDSS optical redshift, and $S_{int}^{25\%\,lim}$ is the flux level at which the completeness of the α.40 catalog falls to 25%, in Jy $\mathrm{km\,s^{-1}}$. According to Eqns. 6 & 7 of Haynes et al. [2011]

$$\log S_{int}^{25\%\,lim} = \begin{cases} 0.5\log W_{50} - 1.312 & \log W_{50}\,, \leqslant 2.5 \\ \log W_{50} - 2.562 & \log W_{50}\,, > 2.5 \end{cases} \tag{3.3}$$

where W_{50} is the HI-line profile width in $\mathrm{km\,s^{-1}}$, measured at the 50% flux level of the profile peak. Since ALFALFA non-detected galaxies lack a measurement of W_{50}, we assign a value based on the average M_*– v_{rot} relation (i.e. the stellar mass Tully-Fisher relation) of α.40 galaxies. We then project the v_{rot} value on the

line-of-sight according to the SDSS r-band axial ratio, and assuming an intrinsic axial ratio of $q_0 = 0.13$ for all galaxies.

Baryonic masses (i.e. stellar mass + atomic gas mass) for all galaxies are calculated as $M_b = M_* + 1.4\,M_{HI}$, where the 1.4 factor is used to account for the cosmic abundance of helium. Note that ALFALFA non-detected galaxies have two assigned values for their HI mass, and consequently two values for their baryonic mass, $M_b^{min} = M_*$ and $M_b^{max} = M_* + 1.4\,M_{HI}^{max}$.

We calculate cumulative mass functions in logarithmic mass bins for all three components (i.e. stellar mass, atomic hydrogen mass, baryonic mass), separately for the HI-selected and optically-selected samples. Since neither sample is volume-limited, mass functions have to take into account the sample selection criteria as well as the large-scale structure in the survey volume. HI selection is based on a combination of galactic HI integrated flux, S_{int}, and profile width, W_{50} [see §3.2.1 & discussion in Section 6 of Haynes et al., 2011]; as a result, galaxies of different HI masses and linewidths are detected out to different distances. Similarly, our optically-selected sample is a flux-limited sample, which results in galaxies with different r-band absolute magnitudes being detected in different volumes. As a result, mass distributions are calculated by summing up the number of detections in a given mass bin (see Fig. 3.2), with each detection weighted by an appropriate volume factor. Individual weighting factors are calculated via the "$1/V_{eff}$" method, as implemented in Zwaan et al. [2005]. This is a non-parametric, maximum-likelihood method, which reduces to the standard $1/V_{max}$ method [Schmidt, 1968] when applied to a spatially homogeneous galactic sample. The advantage of $1/V_{eff}$ consists in the fact that it is insensitive to local density fluctuations, and hence mostly immune to structure-induced bias.

Full details of the method definition and implementation can be found in the Appendix.

Lastly, a fraction of galaxies that satisfy all criteria for SDSS spectroscopic followup cannot be observed for technical reasons (mostly fiber collisions), and are therefore not included in the SDSS spectroscopic database. We therefore correct the normalization of all optically-selected distributions by $1/ < f_{spec} >$, using the average spectroscopic completeness value reported in Li and White [2009], $< f_{spec} >= 0.92$. Similarly, a fraction of the ALFALFA volume is "lost" due to RFI contamination of certain frequency bands in the ALFALFA passband. We correct the normalization of all HI-selected distributions by $1/(1 - f_{RFI})$, where $f_{RFI} = 0.03$. We would also like to note that, due to the $4'$ beam size of the ALFA receiver, a number of HI sources are expected to be blended. We do not attempt to correct for blending but, given that HI-selected galaxies are a weakly clustered population [e.g. Martin et al., 2012], we anticipate the effect on the HI-selected distributions to be small.

3.3.2 Results

Figure 3.3 shows the cumulative distribution[3] of stellar mass (SMF; *gold symbols*), HI mass (HIMF; *cyan symbols*) and baryonic mass (BMF; *black symbols*), derived from the HI-selected galaxy sample (values listed in Table **??**). The HI-selected BMF follows closely the HI-selected SMF at high masses, while at the low-mass end the contribution of the HIMF becomes dominant; this is because HI-selected galaxies become more gas-rich as their stellar mass decreases. Figure 3.4 shows the same distributions (SMF, *gold line*; HIMF, *cyan lines*; BMF, *black*

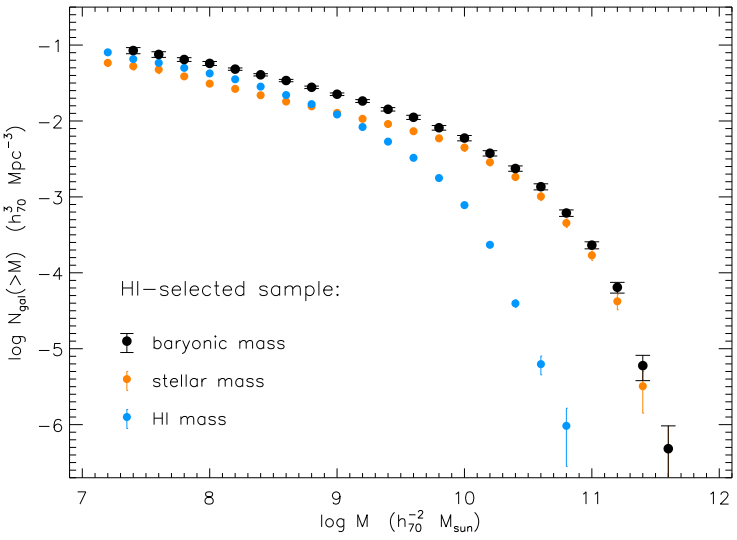

Figure 3.3 The *cumulative* distributions of stellar mass (SMF, *gold symbols*), atomic hydrogen mass (HIMF, *cyan symbols*) and baryonic mass calculated as $M_b = M_* + 1.4\,M_{HI}$ (BMF, *black symbols*), derived from the HI-selected galaxy sample. Error bars represent just the Poisson counting error assuming independent errors among different mass bins.

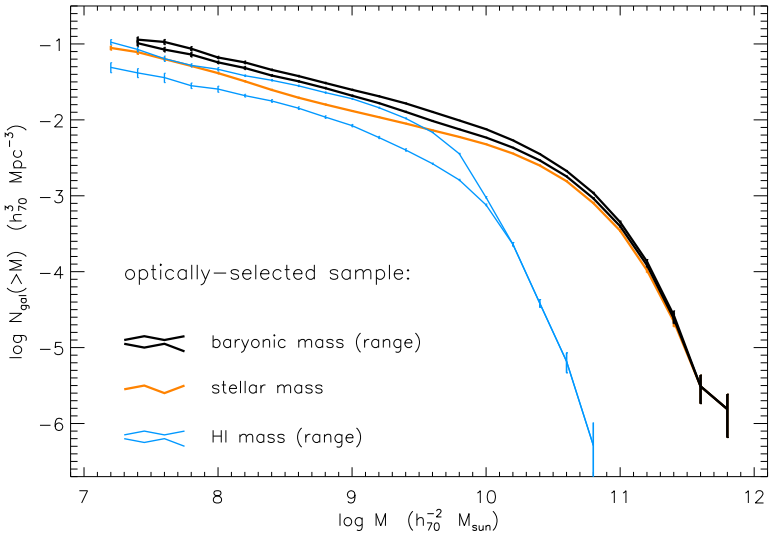

Figure 3.4 The *cumulative* distribution of stellar mass (*gold line*), atomic hydrogen mass (*cyan lines*) and baryonic mass calculated as $M_b = M_* + 1.4\ M_{HI}$ (*black lines*), derived from the optically-selected galaxy sample. The atomic hydrogen and baryonic distributions are represented as allowed ranges, based on estimates of the minimum and maximum HI mass for galaxies that are not detected by ALFALFA ($\{M_{HI}^{min}, M_{HI}^{max}\}$, see §3.2.3 for details). Error bars again represent just the Poisson counting error assuming independent errors among different mass bins.

lines) derived from the optically-selected sample (values listed in Table **??**). Recall that, in the case of the optically-selected sample, a lower and upper limit of the HIMF and BMF are shown, since SDSS galaxies that are undetected by ALFALFA are assigned an upper and lower limit on their HI content, and therefore also on their baryonic mass (see §3.2.3).

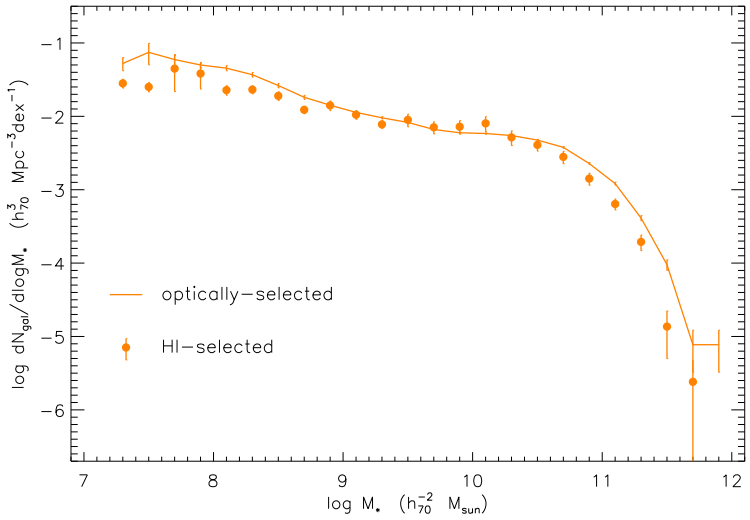

Figure 3.5 Comparison of the differential galactic stellar mass function (SMF) derived from the HI-selected (*gold symbols*) and optically-selected (*gold line*) galaxy samples. Error bars represent just the Poisson counting error on individual mass bins. The optically-selected SMF is systematically higher than the HI-selected SMF at the high-mass ($M_* \gtrsim 10^{11}\,M_\odot$) and low-mass ($M_* \lesssim 10^{8.5}\,M_\odot$) ends. This difference is mostly due to the bias of the HI-selected sample against the red galaxy population (see §4.3 for a detailed discussion).

In Figure 3.5 we compare the SMFs derived from the optically-selected and HI-selected samples. The two SMFs are consistent at intermediate masses, but the optically-selected SMF (gold line) is systematically higher at the high-mass and low-mass ends. At high masses the discrepancy is due to the bias of the HI-selected sample against the most massive galaxies, which are usually red

[3]We show cumulative mass functions in Figs. 3.3 & 3.4, because the cumulative -and not the differential- distributions are directly related to the stellar and baryon galactic fractions computed in Sec. 3.5. All other figures however show differential mass functions, which best display the details and errors of the distributions.

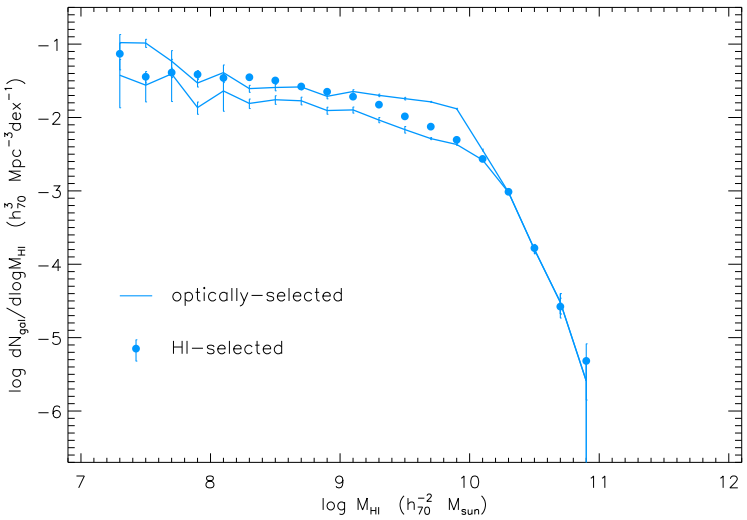

Figure 3.6 Comparison of the differential HI mass function (HIMF) derived from the HI-selected (*cyan symbols*) and optically-selected (*cyan lines*) galaxy samples. Error bars represent just the Poisson counting error on individual mass bins. The HIMFs derived from the two samples are mostly consistent, with the HI-selected HIMF having a slightly steeper low-mass end slope than that suggested by the optically-selected HIMF range. See §4.3 for a detailed discussion.

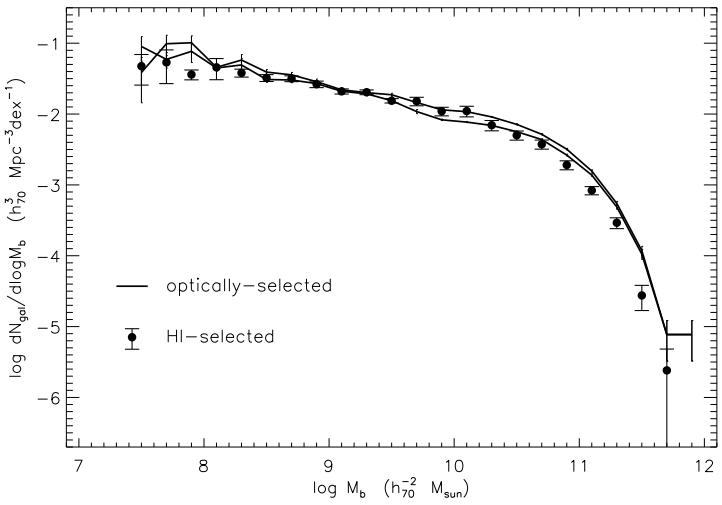

Figure 3.7 Comparison of the differential baryonic mass function (BMF) derived from the HI-selected (*black symbols*) and optically-selected (*black lines*) galaxy samples. Error bars represent just the Poisson counting error on individual mass bins. The optically-selected BMF is mostly consistent with the HI-selected BMF, except at the high-mass end. This is a direct result of the discrepancy between the optically-selected and HI-selected SMFs at high masses (Fig. 3.5). See §4.3 for a detailed discussion.

passive systems. The discrepancy at the low-mass end is mostly due to the population of red-sequence dwarf galaxies in the nearby Virgo cluster[4] that are undetected by ALFALFA; these are mostly dwarfs with early-type morphologies and very low HI content [see Hallenbeck et al., 2012, for example]. On the other hand, the HI-selected and optically-selected HIMFs (Fig. 3.6) are in good agreement with one another, with the HI-selected HIMF having a slightly steeper low-mass end slope than what suggested by the range of the optically-selected distribution. The two BMFs are mostly in agreement with one another, except at the high-mass end (a factor of ≈4 at $M_b = 10^{11.5}\,M_\odot$). This is a direct consequence of the discrepancy between the optically-selected and HI-selected stellar mass functions at high masses. Note that there is little difference between the two BMFs at low masses, which suggests that the low-mass end of the BMF has been measured robustly.

3.3.3 Comparison with other work

Figure 3.8 compares the optically-selected SMF presented in this work (same as gold line in Fig. 3.5) with the local-universe SMF of Baldry et al. [2008] and the Yang et al. [2009] SMF, which are both based on the New York University Value-Added Galaxy Catalog [NYU-VAGC; Blanton et al., 2005b][5]. There is excellent agreement between the Baldry et al. SMF and our optically-selected SMF, especially at intermediate and low stellar masses ($M_* \lesssim 10^{11} M_\odot$). The deviations at high masses are due to the fact that stellar masses in Baldry et al. are calcu-

[4]The presence of a massive cluster (Virgo cluster) at a distance of just 16.5 Mpc from the observer makes the ALFALFA survey volume different from an average cosmological volume. The effect of the presence of the Virgo cluster on the HIMF has been investigated by Martin et al. [2010, §6.1], who found however only minor effects. More generally, the issue of cosmic variance regarding ALFALFA statistical distributions is discussed and quantified in Papastergis et al. [2011, §4.3].

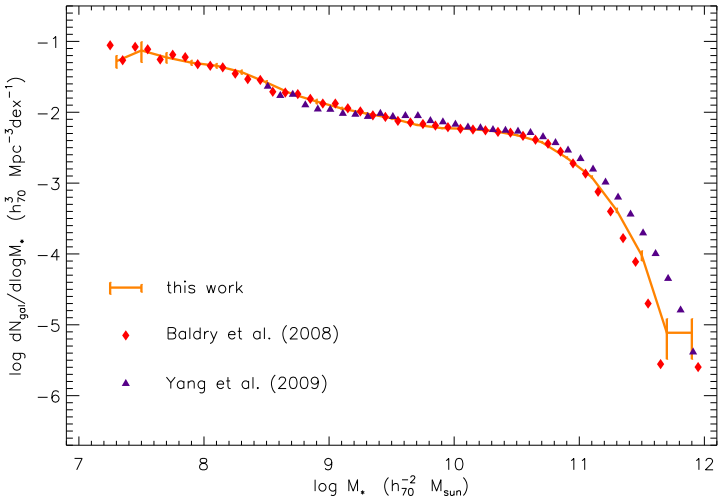

Figure 3.8 The gold solid line with errorbars represents the differential SMF derived in this work from the optically-selected sample (same as gold line in Fig. 3.5). The red diamonds correspond to the Baldry et al. [2008] SMF in the local universe ($z < 0.06$), while the purple triangles correspond to the SMF of Yang et al. [2009], extracted over a larger volume ($z < 0.2$). Both the Baldry at el. and Yang et al. SMFs are based on the NYU-VAGC galaxy catalog. See §3.3.3 for a discussion of the comparison.

lated differently than in this work [for details see Sec. 3 of Baldry et al., 2008]; note that a systematic difference of just 0.1 dex (26%) in stellar mass would be enough to explain the observed difference in abundance at the high-mass end.

The Yang et al. [2009] SMF is systematically higher than our optically-selected SMF at high masses, and displays a more pronounced "plateau" at intermediate masses. It is not clear what the cause of the difference at the high-mass end of the distributions is, but it may relate to the fact that the Yang et al. SMF is extracted from a significantly larger volume than our measurement (the maximum redshift is $z = 0.2$ for the Yang et al. sample and $z = 0.05$ for the sample used in this work). Moreover, Yang et al. use the prescription of Bell et al. [2003] to estimate stellar masses, which is based on the galactic $g - r$ color. As discussed in more detail in §3.4.1, the use of different stellar mass estimators

[5] `http://sdss.physics.nyu.edu/vagc/`

can significantly affect the shape of the measured SMF.

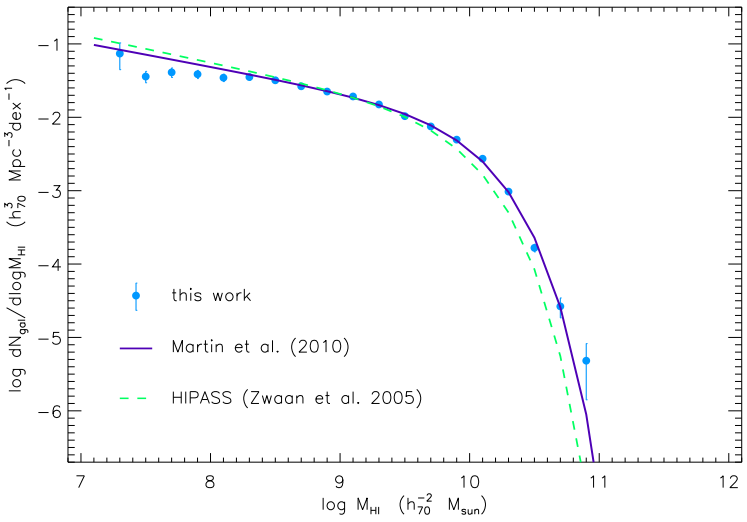

Figure 3.9 The cyan symbols with error bars represent the differential HIMF derived in this work from the HI-selected sample (same as cyan symbols in Fig. 3.6). The solid purple line corresponds to the Schechter function fit to the Martin et al. [2010] HIMF, which is based on the full $\alpha.40$ catalog of ALFALFA sources and without any optical selection cuts. The green dashed line corresponds to the Schechter fit to the Zwaan et al. [2005] HIMF, based on $4\,315$ galaxies detected by the HIPASS survey. See §3.3.3 for a discussion of the comparison.

Figure 3.9 compares the HI-selected HIMF presented in this work (same as the cyan symbols in Fig. 3.6) with the HIMF of Martin et al. [2010] derived from $10\,119$ galaxies detected by ALFALFA over $\approx 2\,600$ deg^2 of sky (purple solid line). There is excellent agreement at intermediate and large HI masses ($M_{HI} \gtrsim 10^{8.5}\,M_{\odot}$) between the Martin et al. [2010] HIMF and the HIMF derived in this work, while at lower masses the Martin et al. [2010] HIMF is slightly steeper than ours. This disagreement may be due to the set of additional optical requirements imposed on our HI-selected sample. As argued in §3.2.1, these requirements are expected to reduce the number of low-mass systems in the sample and therefore decrease the inferred space density at the low-mass end. The dashed green line represents the HIMF of Zwaan et al. [2005] based

on $4\,315$ sources detected by the HI Parkes All-Sky Survey (HIPASS) over the whole southern celestial hemisphere ($\approx 29\,000$ deg^2). There is disagreement between both ALFALFA-based HIMFs and the HIPASS-based HIMF of Zwaan et al. [2005], at the high-mass end. As argued in Martin et al. [2010], the higher sensitivity of the ALFALFA survey compared to HIPASS, which enables ALFALFA to detect HI-massive systems over a larger volume, should give a statistical advantage to the ALFALFA survey in determining the high-mass end of the HIMF. However, the difference is too large to be explained by counting statistics or cosmic variance (e.g. according to the estimates of Somerville et al., 2004 or Driver and Robotham, 2010). On the other hand, due to the exponential drop-off of the HIMF at high masses, a flux calibration difference of as low as 0.1 dex could give rise to a similar discrepancy.

3.4 Uncertainties & systematics

3.4.1 Stellar mass estimator

A variety of methods exist to estimate stellar masses from spectra or broadband photometric measurements of galaxies [Kauffmann et al., 2003, Bell et al., 2003, Brinchmann et al., 2004, Glazebrook et al., 2004, Gallazzi et al., 2005, Panter et al., 2007, Salim et al., 2007, to name a few]. Most methods rely on comparing the actual galactic emission to the light output of a set of galactic stellar population models. The models that most closely reproduce the observed data are then used to estimate the galactic properties of interest (e.g. stellar mass, present star formation rate, internal extinction etc.); it is therefore very important to consider

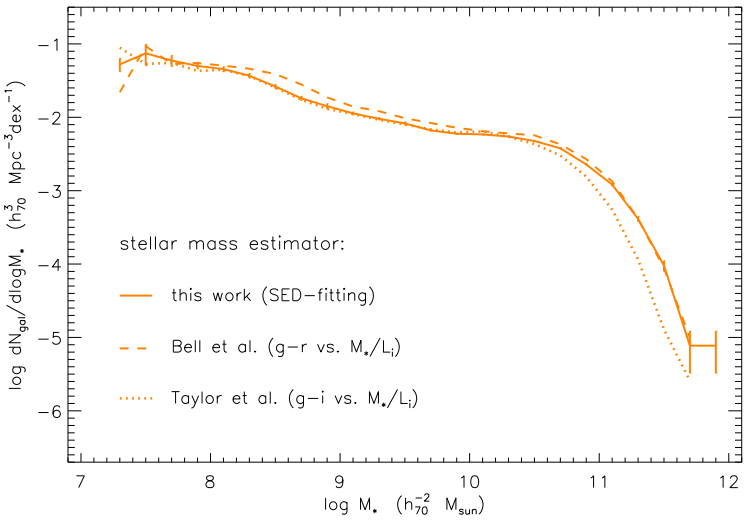

Figure 3.10 The gold solid line with error bars represents the differential SMF derived from the optically-selected sample in this work, using stellar masses based on SED-fitting [Huang et al., 2012b]. The gold dashed line represents the SMF computed from the same sample but using stellar masses derived from the galactic $g-r$ color and the i-band luminosity, according to the widely used Bell et al. [2003] calibration. The gold dotted line represents the SMF computed using stellar masses derived from the galactic $g-i$ color and the i-band luminosity, according to the the more recent calibration of Taylor et al. [2011]. See §3.4.1 for a discussion of the comparison.

model stellar populations which span as large a range of physical parameters as the galaxies in the sample being studied. Inferred galactic properties depend not only on the particular type of data employed by each method (e.g. spectroscopy vs. broadband photometry or optical vs. near-infrared photometry), but also by differences in the way in which the model stellar populations are constructed. This means that different methods can yield different estimates of a galactic property even when the *same* observational measurements are used. For example, Pforr et al. [2012] find that unbiased stellar masses can only be recovered when the true star formation history (SFH) of a galaxy is known. In practice however a restricted set of SFHs is considered (often in the form of a parametrized function), which may introduce systematics for galaxies with SFHs that are not well described by the assumed general form. Additional complications can be introduced by the different treatment of dust reddening among different models. In general, stellar mass estimates can differ systematically by as much as 0.3 dex, while for individual galaxies the scatter can be as large as 0.6 dex [Pforr et al., 2012].

Here we compare the optically-selected SMF presented in this work (computed from stellar masses derived from SED-fitting, see §3.2.3) with the SMF obtained using stellar masses derived from a single galactic color, according to the widely-used Bell et al. [2003] calibration as well as the more recent calibration of Taylor et al. [2011]. More specifically, we compute Bell et al. masses by multiplying the i-band luminosity of each galaxy by a mass-to-light ratio inferred from its $g-r$ color[6]. We choose this particular combination of bands because it is relatively immune to contamination of galactic spectra by bright nebular emission lines [West et al., 2009]. We use a similar procedure to calculate Taylor et al. masses, by using their calibration of i-band mass-to-light ratio

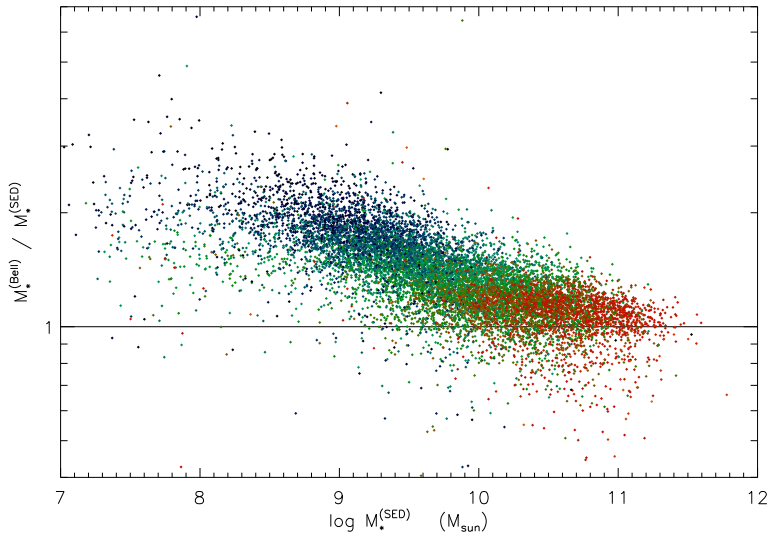

Figure 3.11 A galaxy-by-galaxy comparison of stellar masses derived from SED fitting of the SDSS u, g, r, i, z bands used in this work (see §3.2.3), and those derived from the galactic $g - r$ color and the i-band luminosity according to the widely used Bell et al. [2003] calibration. Each datapoint corresponds to a galaxy in the optically-selected sample, and the symbol color represents the g-r color of the galaxy. The two stellar mass estimates agree fairly well for red passive galaxies, while for blue star-forming galaxies Bell et al. masses are systematically larger by up to a factor of ≈2 (see §3.4.1 for a detailed discussion).

versus $g-i$ color. Taylor et al. [2011] argue that using $g-i$ colors best constrains the galactic stellar mass estimates.

Figure 3.10 shows the impact that different methods of estimating stellar mass have on the measurement of the SMF. When Bell et al. stellar masses are used (gold dashed line), the SMF becomes systematically higher at low and intermediate masses ($M_* \lesssim 10^{11}\,M_\odot$), while it remains mostly unchanged at the high-mass end. The reason for this pattern becomes evident in Fig. 3.11, where we see that Bell et al. masses agree fairly well with the masses derived in this work for red passive galaxies, but are systematically larger (by up to a factor of $\approx$2) for blue star-forming galaxies. Huang et al. [2012b] argue that the difference can be primarily attributed to the fact that the stellar population models used for the Bell et al. calibration do not consider "bursty" star formation histories which are typical of low-mass galaxies with blue colors. This leads to systematically older stellar populations for blue galaxies according to the Bell et al. method, which in turn results in systematically higher stellar mass estimates. Note that including models with bursty SFHs in a stellar population library does not by itself guarantee a correct estimate of stellar mass; overestimating the effect of bursts would result in systematically low stellar masses for blue galaxies. Conversely, when Taylor et al. masses are used (gold dotted line), the SMF becomes systematically lower at the high-mass end, while it is mostly unchanged at low and intermediate masses. Again, this is a result of the fact that Taylor et al. masses agree well with the SED-fitting masses used in this work for blue star-forming galaxies, but are systematically lower (by up to a factor of $\approx$1.4) for red passive galaxies.

[6]We use SDSS colors, computed from Galactic extinction-corrected model magnitudes (`modelmag`), to calculate mass-to-light ratios in the i-band. i-band luminosities are then calculated from the i-band Petrosian magnitudes reported in SDSS (`petromag`), corrected for Galactic extinction according to the values listed in the SDSS database. The solar absolute magnitude

3.4.2 Distance uncertainties

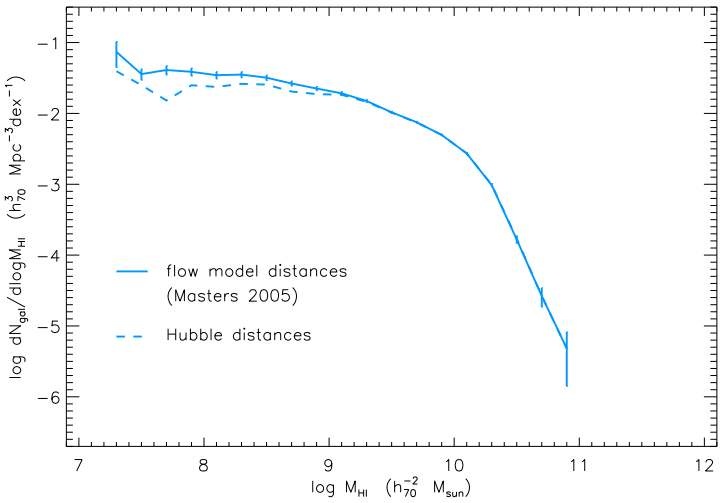

Figure 3.12 The cyan solid line with error bars represents the differential HIMF derived from the HI-selected sample in this work, using flow model distances [Masters, 2005] for most nearby galaxies. The cyan dashed line represents the HIMF computed from the same sample but using simple Hubble distances for all galaxies. Distance uncertainties affect primarily the mass estimates of nearby galaxies and so the main effect is a change of the low-mass end slope of the distribution (see §3.4.2 for further discussion).

Stellar, HI and baryonic masses are distance-dependent quantities and hence their statistical distributions are affected by distance uncertainties. This is particularly true for the low-mass end of the distributions, which is determined by the properties of nearby galaxies; neglecting the peculiar velocity of some of these objects can cause fractional distance and mass errors of order $\approx 100\%$, especially in a volume with complex large-scale structure such as the one surveyed by ALFALFA (see Fig. 3.1).

For this reason, nearby galaxies ($v_{CMB} \leqslant 6\,000$ km s^{-1}) in the α.40 catalog are assigned distances based on a parametric peculiar velocity flow model [Masters, 2005], and only more distant galaxies ($v_{CMB} > 6\,000$ km s^{-1}) are assigned simple Hubble distances according to their CMB recessional velocity. The Masters

in the i-band is taken to be $M_{\odot,i} = 4.57$.

[2005] flow model includes a dipole and a quadrupole component (local group bulk motion & asymmetric expansion) and two local attractors (Virgo cluster & Great Attractor), and is calibrated against the SFI++ catalog of galaxies with redshift and independent distance distance information (from Tully-Fisher). The residuals are then attributed to random thermal motions, which are estimated to have a magnitude of $\sigma_{local} = 160\,\mathrm{km\,s^{-1}}$. In addition, distances reported in the α.40 catalog take into account known group and cluster membership as well as primary distance information published in the literature. This latter information is not available for the majority of the galaxies in our optically-selected sample (which are not included in α.40), and we only make an attempt to assign all probable Virgo members to the Virgo cluster distance ($D = 16.5\,\mathrm{Mpc}$).

Here we re-evaluate the HIMF for our HI-selected sample using uniformly Hubble distances for all galaxies, in order to illustrate the impact of the distance assignment scheme on the derived distributions. Figure 3.12 shows that the HIMF computed using Hubble distances (cyan dashed line) has a much shallower low-mass end slope compared to the HIMF presented in this work, which uses flow model distances for most nearby galaxies (cyan solid line). This result is in agreement with the work of Masters et al. [2004], who find that neglecting the local peculiar velocity field when estimating distances to nearby galaxies in the ALFALFA volume will lead to a systematically shallower low-mass end slope for the HIMF.

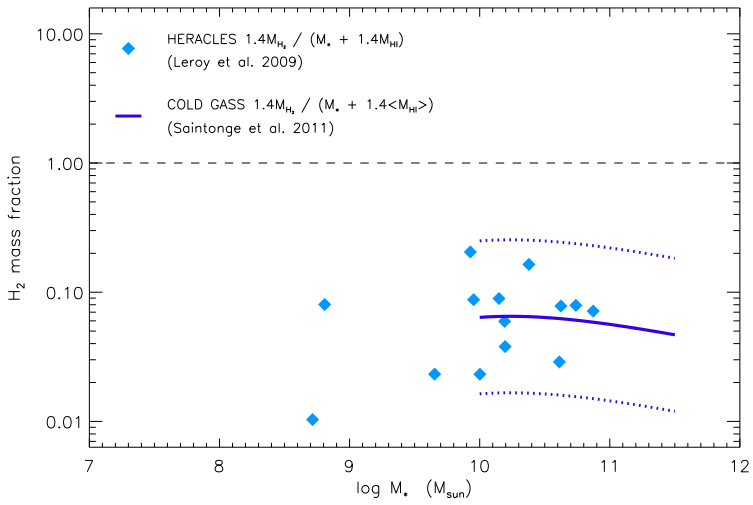

Figure 3.13 Ratio of molecular gas mass (corrected for the abundance of helium) to the "baryonic mass" computed as $M_b = M_* + 1.4\,M_{HI}$. Cyan diamonds represent 14 galaxies of the HERACLES survey [Leroy et al., 2009] detected in CO line emission. The solid and dotted blue lines represent the average and 2σ scatter of the distribution found for 125 CO detected galaxies in the COLD GASS survey [Saintonge et al., 2011]. Both samples show that, at least for $M_* > 10^{8.5} M_\odot$, molecular gas is almost always a subdominant mass component (see §3.4.3 for details).

3.4.3 Molecular & ionized gas

Throughout this chapter we use the term "baryonic mass" to refer to the sum of the stellar and atomic gas mass ($M_b = M_* + 1.4\,M_{HI}$), a convention that is common in the literature. This definition however excludes a number of baryonic components that are definitely present in galaxies, most notably molecular and ionized -warm or hot- gas.

Figure 3.13 displays the fraction of molecular hydrogen (H_2) mass (accounting for helium) to "baryonic mass" as defined above, as a function of stellar mass. The cyan diamonds represent 14 galaxies from the HERACLES survey [Leroy et al., 2009] with H_2 masses measured from interferometric CO line observations (using a fixed α_{CO} conversion factor). Over the probed stellar mass

range ($M_* \gtrsim 10^{8.5} M_\odot$), molecular gas rarely contributes more than 10% of the "baryonic mass". The same conclusion is reached when H_2 mass measurements from the COLD GASS survey are considered [Saintonge et al., 2011]. The blue solid and dotted lines show the mean and 2σ scatter of the relation between the molecular and "baryonic" mass components, based on 125 galaxies detected in CO emission with the IRAM 30m telescope. In this latter case we have estimated the atomic hydrogen mass of galaxies indirectly, using the average M_{HI}/M_* vs. M_* relation of the COLD GASS parent sample [Catinella et al., 2010]. Again, over the stellar mass range probed by the survey ($M_* = 10^{10} - 10^{11.5}\, M_\odot$) molecular gas is always a subdominant mass component. At lower stellar masses there is large uncertainty on the fractional contribution of H_2, as it is not precisely known how well the galactic CO emission traces molecular hydrogen mass. In particular, the α_{CO} conversion factor may vary by about an order of magnitude as we consider less luminous and more metal poor late-type galaxies [e.g. Boselli et al., 2002].

Determining the contribution of ionized gas to the total baryonic mass budget of galaxies is much more challenging. For example Reynolds [2004] argue that warm ionized hydrogen (HII) may amount to about 1/3 of the mass of atomic hydrogen (HI) in the disk of the Milky Way. If the ratio of ionized-to-neutral hydrogen mass ($f_{\frac{HII}{HI}}$) were fairly constant among galaxies, then the baryonic mass of a galaxy would be given by the expression $M_b = M_* + 1.4\,(1 + f_{\frac{HII}{HI}}) M_{HI}$. If $f_{\frac{HII}{HI}} \approx 0.3$, then the peak value of the η_b - M_h relation (see Fig. 3.17) would shift to lower halo mass and the peak value would slightly increase. However, since the precise value and scatter of $f_{\frac{HII}{HI}}$ is not well constrained -and its dependence on galaxy size is not known- we choose not to include the contribution of warm ionized gas in the calculation of M_b.

Assessing the contribution of the hot ionized medium (HIM) to the total baryonic mass of a galaxy is even more challenging. The coronal HIM may be the dominant baryonic mass component in galactic halos, especially in massive ellipticals. Determining the mass contribution of the HIM for less massive galaxies however is observationally challenging. In any case, the tightness of the "baryonic Tully-Fisher relation" when computed just from the stellar and HI mass [e.g. McGaugh, 2012, Hall et al., 2012] implies that the HIM never dominates the total baryonic mass budget of late-type galaxies, at least within the extent of the galactic HI disk.

3.5 The stellar & gas content of DM haloes

3.5.1 The abundance matching method and its application

Let $N_{\rm gal}(M_b)$ be the cosmic number density of galaxies with baryon mass greater than M_b and let $N_h(M_h)$ be the cosmic number density of haloes with mass greater than M_h. The fundamental assumption of the abundance matching method [Marinoni and Hudson, 2002; Vale and Ostriker, 2004; see also Behroozi et al., 2010 for a review] is that M_b is a monotonically increasing function of M_h. With this assumption, $M_b\,(M_h)$ can be determined by solving the equation

$$N_{\rm gal}(M_b) = N_h(M_h). \tag{3.4}$$

In reality, the baryon content of a halo will depend not only on its mass but also on other parameters, such as its formation history. As a result, a scatter

in the distribution of M_b at a given M_h is expected. Neglecting the scatter is, nevertheless, justifiable because the aim of abundance matching is precisely to determine the average value of M_b within a halo of mass M_h.

We evaluate the right hand side of equation 3.4, using a halo mass function extracted from one of the cosmological N-body simulations of the Horizon Project[7]. The simulation was run with a public version of the GADGET code [Springel et al., 2005], and uses 1024^3 particles of mass $m_p \sim 8.5\ 10^7\ M_\odot$ to simulate the formation and evolution of DM structures in a comoving volume of 100 h^{-1} Mpc on a side. It assumes a cosmology and initial conditions which are consistent with Wilkinson Microwave Anisotropy Probe (WMAP) third year results [Spergel et al., 2007], namely $h = 0.73$, $\Omega_\Lambda = 0.76$, $\Omega_m = 0.24$, and $\sigma_8 = 0.76$. The identification of DM haloes and sub-haloes was done with the adaptaHOP algorithm presented in Aubert et al. [2004]. Haloes are identified as groups of particles above a threshold over-density of 80 times the mean density of the universe, which corresponds to a mean overdensity contrast of about 200. The identification of subhaloes within haloes is done using the method described in Tweed et al. [2009]. We only keep haloes and sub-haloes with more than 20 particles, i.e. we introduce a minimum halo mass of $M_h \approx 1.7\ 10^9\ M_\odot$.

An important issue for our analysis is determining whether subhaloes should be included or excluded in the calculation of the HMF used in the abundance matching procedure. Let $M_*^c(M_h)$ be the stellar mass of the central galaxy in a halo of mass M_h and let $M_*^s(M_h)$ be the stellar mass of the satellite galaxy in a subhalo of mass M_h. If subhaloes are excluded, then we implicitly assume that all galaxies in our samples are central galaxies, that is, $M_*^s(M_h) = 0$. On the other hand, if subhaloes are included, we implicitly assume that the M_b - M_h relation

[7] `http://www.projet-horizon.fr`

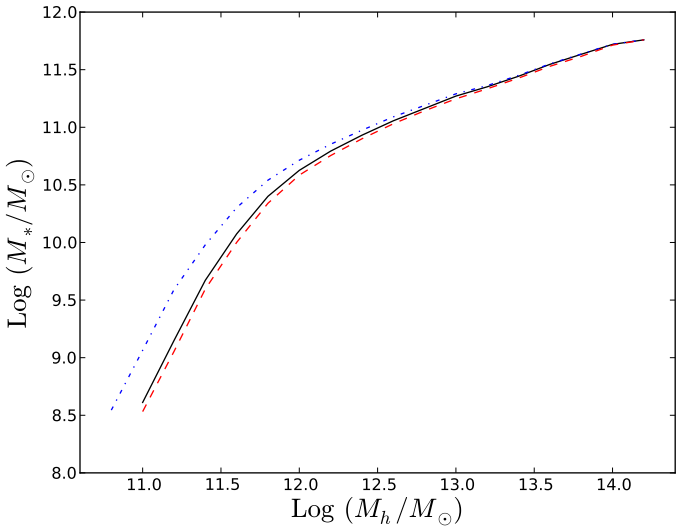

Figure 3.14 Effects of the choice of stellar and halo mass functions on the M_* - M_h relation. The black solid line shows the M_* - M_h relation obtained by matching the stellar mass function of central galaxies to the mass function of haloes excluding subhaloes. This matching should reproduce the "true" relation for central galaxies. The blue dot-dashed line is the relation obtained by matching the total galaxy stellar mass function, including central and satellite galaxies, to the mass function of haloes excluding subhaloes. The red dashed line is the relation obtained by matching the total galaxy stellar mass function with the mass function of haloes including subhaloes. Note that, unlike all other figures, this figure uses the stellar mass functions of Yang et al. [2009], who have separately measured the SMF for central and satellite galaxies.

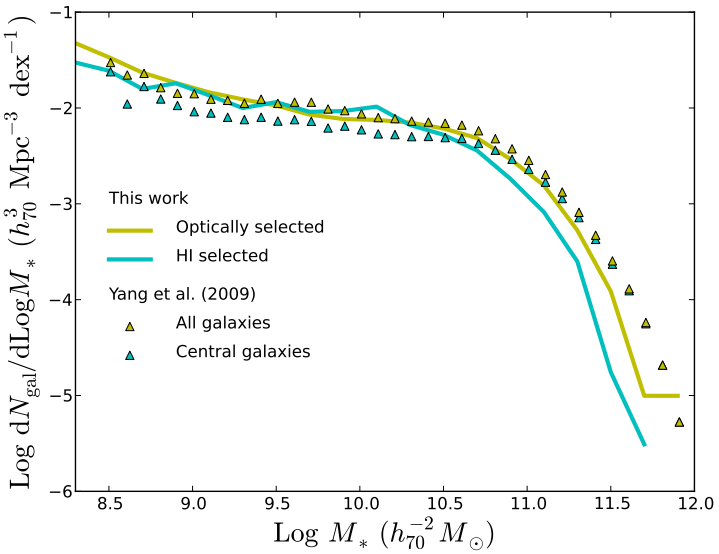

Figure 3.15 The differential SMFs derived from our optically selected sample (yellow curve) and our HI-selected sample (cyan curve) compared to the results of Yang et al. [2009] for the total SMF (yellow symbols) and the SMF for central galaxies only (cyan symbols).

is the same for both central and satellite galaxies, that is, $M_*^s(M_h) = M_*^c(M_h)$.

In order to understand how much this choice could affect our results, we consider the stellar mass functions of Yang et al. [2009], for which separate distributions for central and satellite galaxies have been presented. This allows us to calculate the M_* - M_h relation in three ways: firstly, we can match the SMF of central galaxies with the halo mass function excluding subhaloes; this method will give us the correct M_* - M_h for central galaxies, shown by the black solid line in figure 3.14. Then, we consider the total SMF of central plus satellite galaxies and match it with the halo mass function excluding subhaloes. This is equivalent to assuming that all galaxies are central and overestimates the M_* - M_h for central galaxies (blue dot-dashed line in Fig. 3.14). Finally, we can match the total SMF of central plus satellite galaxies with the total halo mass function including subhaloes. The result, shown by the red dashed line in Figure 3.14,

lies below the black solid line because it is effectively a weighted average of the relation for the dominant central galaxy population (the black solid line) and the relation for the satellite population, which has lower M_* for a given M_h. Quantitative comparisons of $M_*^s(M_h)$ and $M_*^c(M_h)$ have received the attention of much recent literature [see e.g. Cattaneo et al. 2012, submitted, Rodríguez-Puebla et al., 2012, Reddick et al., 2012].

As we do not distinguish between central and satellite galaxies in the samples used in this work, we shall choose the third method (the one that corresponds to the red dashed line) as our best estimator of the M_* - M_h relation and the M_b - M_h relation. This choice will introduce some systematic bias which is, nonetheless, smaller than the typical uncertainty involved in the determination of the M_* - M_h relation.

We also considered whether abundance matching our HI-selected SMF with the HMF excluding subhaloes would give consistent results with our fiducial abundance matching result, obtained by matching our optically-selected SMF and the HMF including subhaloes. Physically, this consideration was motivated by the fact that satellite galaxies tend to be redder than central galaxies of the same mass [e.g. Weinmann et al., 2006], and so HI-selection may be equivalent to the exclusion of satellite galaxies from an observational sample. However, the comparison in Fig. 3.15 of our HI-selected SMF with the SMF for central galaxies of Yang et al. [2009] shows that this argument may not be valid. We note that the comparison between the Yang et al. SMF for central galaxies and our HI-selected SMF is subject to observational systematics, such as the distance assignment scheme or the stellar mass estimator; for example using Hubble distances for the galaxies in our HI-selected sample (see Fig. 3.12) would bring the

two distributions into fair agreement.

3.5.2 Results

Our abundance matching analysis produces two main results:

1. We determine M_*/M_h as a function of M_h by comparing the galaxy stellar mass function (SMF) from our optical sample to the total halo mass function (including subhaloes).
2. We determine M_b/M_h as a function of M_h by comparing the baryonic mass function (BMF) from our optically selected sample to the total halo mass function (including subhaloes).

The latter relation is the focus of this chapter, but we first discuss the former because the results can be compared to an extensive literature of previous studies. The consistency of our findings with previous work on point 1 boosts our confidence that our conclusions on point 2 are robust.

Figure 3.16 shows our result for M_*/M_h as a function of M_h, obtained from our optically selected sample (gold thick solid line). Our analysis extends to halo masses as low as $M_h \approx 10^{10.5}\,M_\odot$, since both our optically-selected and HI-selected samples probe masses down to $M_* \approx 10^7\,M_\odot$. At the high mass end, our relation stops at $M_h \approx 10^{14}\,M_\odot$ because our galactic samples are drawn from a relatively small volume, and are not appropriate to measure the abundance of the most massive galaxies and clusters. We find good agreement with previous estimates from abundance matching obtained in the same manner [Moster et al.,

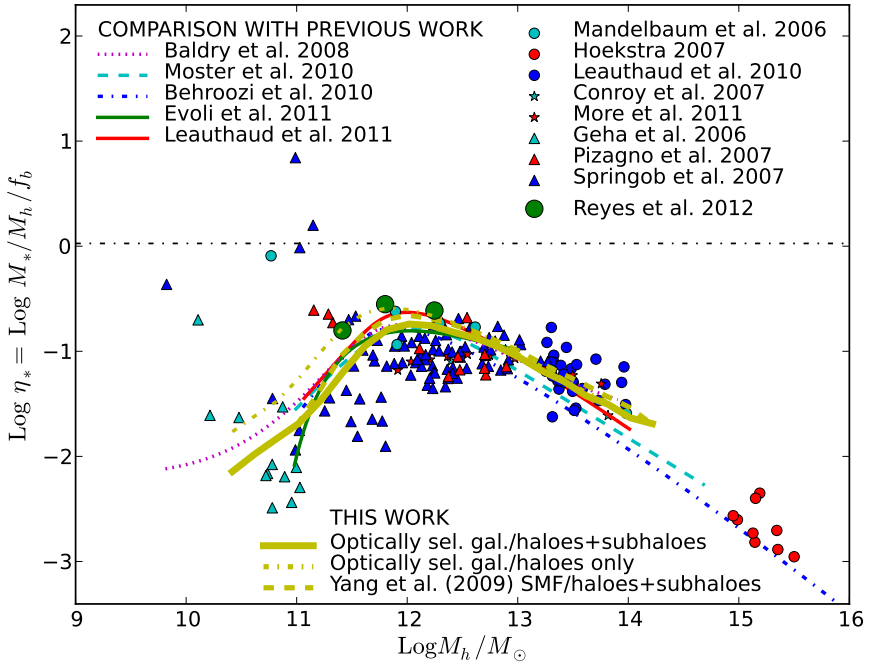

Figure 3.16 The ratio of galactic stellar mass to halo mass as a function of host halo mass ($M_*/M_h - M_h$ relation). The thick yellow line shows our main result, obtained from abundance matching the stellar mass function of our optically-selected sample with the halo mass function including subhaloes. The yellow dashed and dash-dotted lines correspond to variations of our main result, obtained by considering the Yang et al. [2009] SMF and excluding subhaloes from the HMF respectively, and are shown to illustrate uncertainties. The magenta dotted, cyan dashed, blue dot-dashed, green solid and red solid lines correspond to the abundance matching results of Baldry et al. [2008], Moster et al. [2010], Behroozi et al. [2010], Evoli et al. [2011] and Leauthaud et al. [2012], respectively. The big green circles are the results of a stacked weak lensing study of SDSS galaxies by Reyes et al. [2012]. All other data points refer to measurements for individual galaxies: the small circles correspond to galaxies with halo mass measurements from weak lensing studies [Mandelbaum et al., 2006: cyan circles; Hoekstra, 2007: red circles; Leauthaud et al., 2010: blue circles]. The star symbols show galaxies in which the halo mass was determined from studies of stellar dynamics [Conroy et al., 2007: cyan stars; More et al., 2011: red stars]. The triangles show galaxies with halo masses determined from the disc rotation speed [Geha et al., 2006: cyan triangles; Pizagno et al., 2007: red triangles; Springob et al., 2007: blue triangles]. The dotted-dashed horizontal line shows the cosmic baryon fraction $f_b \approx 0.16$.

2010: cyan dashed line; Behroozi et al., 2010: blue dotted-dashed line; Evoli et al., 2011: green solid line; Leauthaud et al., 2012: red solid line; Baldry et al., 2008[8]: purple dotted line]. The thick yellow dash-dotted and dashed lines are shown to illustrate the systematics introduced by the choice of HMF and SMF: the former represents the abundance matching result when our fiducial SMF is matched the HMF excluding subhaloes; the latter is the result of matching the Yang et al. SMF with our fiducial HMF, which includes subhaloes.

We also compare our average M_* - M_h relation with values measured for individual galaxies. The small cyan, red and blue circles correspond to galaxies with measurements of their halo mass M_h from weak lensing [Mandelbaum et al., 2006, Hoekstra, 2007, Leauthaud et al., 2010]. The star symbols correspond to galaxies for which M_h was estimated from stellar dynamics [Conroy et al., 2007, More et al., 2011]. The triangles correspond to disc galaxies for which M_h was determined from their rotation speed [Geha et al., 2006, Pizagno et al., 2007, Springob et al., 2007]. We remark that, while results for individual galaxies have large scatter, they seem to be systematically lower than any of the relations inferred from abundance matching (at least over the halo mass range $M_h = 10^{11} - 10^{12}\,M_\odot$). Furthermore, the halo mass for which M_*/M_h has a maximum appears to be higher when inferred from measurements of individual galaxies compared to the value derived from abundance matching: in the first case the peak is at $M_h \approx 10^{12.5}\,M_\odot$, while in the second case it is at $M_h \approx 10^{12}\,M_\odot$. There is also slight disagreement of all abundance matching results with the results of Reyes et al. [2012] (large green circles), who used a stacked weak lensing analysis of over a hundred thousand disk galaxies in SDSS separated in three bins of stellar mass. Regardless of the method used, however, there is a clear

[8]The Baldry et al. [2008] abundance matching result uses a "galactic" halo mass function by Shankar et al. [2006].

consensus that M_*/M_h is much lower than the universal baryon fraction $f_b \approx 0.16$ (horizontal black dash-dotted line in figure 3.16), at all halo masses.

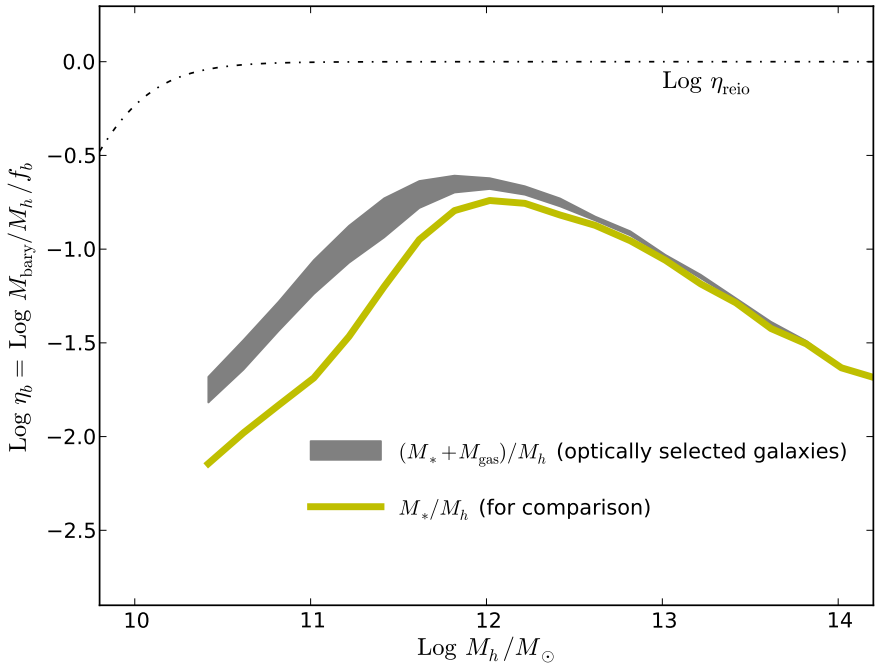

Figure 3.17 The baryon fraction of galaxies, including stars and atomic gas, as a function of their host halo mass ($M_b/M_h - M_h$ relation). The gray shaded area shows the results of an abundance matching analysis of our optically-selected sample. Its boundaries correspond to two extreme assumptions for the gas content of galaxies detected optically but not in HI: i) galaxies that are not detected in HI contain no gas (lower boundary) and ii) galaxies that are not detected in HI contain the largest amount of gas that could have escaped detection from ALFALFA (upper boundary). The thick solid yellow line is the M_*/M_h-M_h relation (same as in figure 3.16), and is shown for comparison. The dotted-dashed line shows the baryon fraction that Okamoto et al. [2008] predict based on hydrodynamic simulations that include cosmic reionisation.

Let us now examine the results for the M_b/M_h - M_h relation (Fig. 3.17). The gray shaded area shows the relation derived from our optically-selected sample, matched to a halo mass function that includes both haloes and subhaloes: its upper and lower envelopes correspond to the distribution for M_b^{max} and M_b^{min}, respectively, as defined in § 3.2.3. The thick gold solid line represents the relation for the stellar mass (same as in Figure 3.16) and has been added for reference.

Figure 3.17 shows that η_b decreases monotonically with decreasing halo mass, despite the fact that atomic gas contributes progressively more to the baryonic mass budget of galaxies.

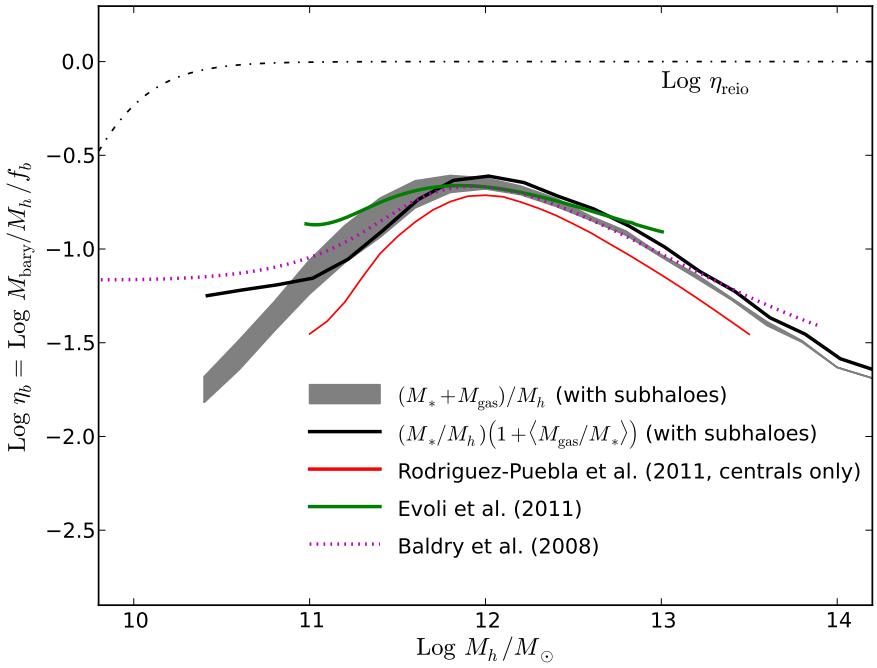

Figure 3.18 The gray shaded area represents the M_b/M_h - M_h relation derived in this work from our optically-selected sample (same as Fig. 3.17). The magenta dotted line and the green solid lines show the M_b/M_h - M_h relations that Baldry et al. [2008] and Evoli et al. [2011] derived from their M_*/M_h - M_h relations (Fig. 3.16), using the mean gas-to-stellar mass ratio as a function of stellar mass to account for the gas content of galaxies. The black solid line correspond to the results obtained from our optically-selected sample when we use the same method. The red solid line represents the result of Rodríguez-Puebla et al. [2011], who used separate f_{HI} - M_* relations for blue and red central galaxies. Lastly, the thin black dotted-dashed line is the same as Fig. 3.17 and shows the baryon fraction that Okamoto et al. [2008] predict based on hydrodynamic simulations that include cosmic reionisation.

In Fig. 3.18 we compare our results to those by Baldry et al. [2008] and Evoli et al. [2011], who derived their M_b/M_h - M_h relations from the equation

$$\frac{M_b}{M_h} = \frac{M_*}{M_h}\left(1 + \langle\frac{M_{gas}}{M_*}\rangle\right). \tag{3.5}$$

In Eqn. 3.5 the baryonic mass is computed from the stellar mass, using the mean gas-to-stellar mass ratio for galaxies as a function of stellar mass. To enable a cleaner comparison, we have made the exercise of re-deriving the M_b/M_h - M_h relation from our M_*/M_h - M_h relation, using the same procedure followed by Baldry et al. [2008] (as in Eq. 3.5). The result is shown by the black solid line in Fig. 3.18. The main difference between the results obtained by using individual galaxy gas masses (gray shaded region) and by adopting a mean gas-to-stellar mass ratio (black solid line) seems to be an artificial flattening of the M_b/M_h relation at low masses. This is probably due to the fact that the latter method ignores the large scatter of galactic M_{HI}/M_* values from the mean, and leads to the incorrect interpretation that the baryon retention fraction (η_b) of low-mass halos asymptotes to some fixed value. The red solid line in Fig. 3.18 corresponds to the result of Rodríguez-Puebla et al. [2011], who used a separate mean M_{HI}/M_* - M_* relation for blue and red galaxies. Their result[9] shows no signs of flattening at low masses, however the relation only extends down to $M_h = 10^{11}\,M_\odot$. Independently of the used method however, all results point to values of η_b that are well below unity, and cannot be explained by the effects of cosmic reionization alone (black dotted-dashed line in Fig. 3.18; see Sec. 3.5.3 for details).

The $\langle M_{gas}/M_* \rangle$ relation that we use to evaluate the right hand side of Eqn. 3.5 (black solid line in Fig. 3.19) is a power-law fit to the M_{HI}/M_* data by Swaters and Balcells [2002], Garnett [2002], Noordermeer et al. [2005] and Zhang et al. [2009]. The best fit relation, plotted as a thin solid black line in Fig. 3.19, is given by $\log(M_{HI}/M_*) = -0.43\,\log(M_*/M_\odot) + 3.75$. The blue contours in Fig. 3.19 represent the distribution of M_{HI}/M_* values for the galaxies in our optically-

[9]Note that the Rodríguez-Puebla et al. [2011] result refers to central galaxies only.

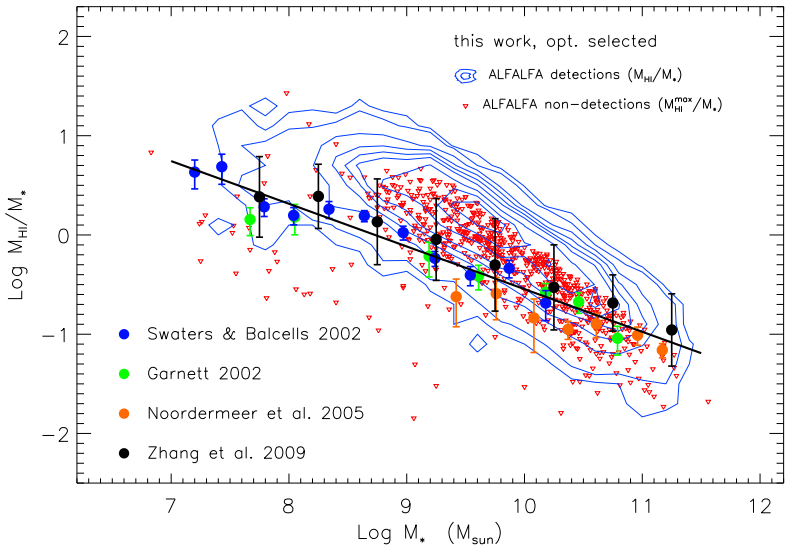

Figure 3.19 HI-to-stellar mass ratio vs. galaxy stellar mass. The symbols with error bars show the results of Swaters and Balcells [2002, blue circles], Garnett [2002, green circles], Noordermeer et al. [2005, orange circles] and Zhang et al. [2009, black circles] for the average HI-to-stellar mass ratio in bins of stellar mass. The black line is a power-law fit to these data points. The blue contours represent the distribution of HI-fraction (M_{HI}/M_*) for the galaxies in our optically-selected sample that are detected by ALFALFA, while the red inverted triangles are maximum HI-fractions (M_{HI}^{max}/M_*) for a representative subset of our optically-selected galaxies that are not detected by ALFALFA.

selected sample that are detected by ALFALFA. Since the ALFALFA survey is a blind HI survey with a fixed integration time per pointing, the ALFALFA distribution is expected to lie above the data obtained by pointed observations of optically-selected galaxies [as in Swaters and Balcells, 2002, Garnett, 2002, Noordermeer et al., 2005]. The inverted red triangles correspond to the maximum HI-fraction (M_{HI}^{max}/M_*) for a representative subsample of our optically-selected galaxies that lack an ALFALFA detection. Note that these upper limits are also systematically higher than the relationship indicated by the power-law fit.

3.5.3 Discussion

The main result of this chapter is the large "gap" between the present-day baryon fraction of galaxies in low-mass halos and the cosmic value ($f_b \approx 0.16$), which is present even when the atomic gas content of the galaxies is taken into account. This result is not contrived, given that atomic gas dominates the baryonic mass budget of galaxies with $M_h \lesssim 10^{11}\,M_\odot$. Moreover, the low-mass behavior is in disagreement with previous studies [e.g. Baldry et al., 2008, Evoli et al., 2011], who find that the η_b - M_h relation flattens out at low masses, and approaches a roughly constant value of $\eta_b \approx 10\%$. These previous results would then require an exceptionally low efficiency of gas-to-stars conversion in low mass systems to explain the observed values of η_*, which decrease monotonically.

In Fig. 3.17 we compare our result for the baryon fraction of halos to the predictions at $z = 0$ of a cosmological hydrodynamic simulation that includes heating from a photoionizing UV background [Okamoto et al., 2008, black dotted-

dashed curve]. At the high-mass end ($M_h \gtrsim 10^{13}\,M_\odot$), the low values of η_b can be explained by the fact that infalling gas is shock-heated [Kereš et al., 2005, Dekel and Birnboim, 2006], and is kept hot by feedback from active galactic nuclei [Croton et al., 2006, Cattaneo et al., 2006, Bower et al., 2006]. This picture is supported by considerable observational evidence in the case of X-ray groups and clusters, but the situation within individual galaxies is not so well understood [see Cattaneo et al., 2009 for a review].

At the low mass end, photoionization heating is expected to become important, since the intergalactic medium is too hot to fall into the shallow potential wells of haloes with $M_h < 10^{10}\,M_\odot$, and their baryon fraction is heavily suppressed. However, this process alone cannot account for the shape of the η_b - M_h relation at low masses, especially for the onset of a sharp decrease in the M_b/M_h ratio at relatively large halo masses ($M_h \approx 10^{11.5}\,M_\odot$). Additional feedback is therefore needed, usually attributed to the ejection of baryons by stellar-driven winds. Semianalytic models of galaxy formation based on this assumption reproduce a good fit to luminosity functions in the local universe [Guo et al., 2011, Benson and Bower, 2010, Somerville et al., 2008, Cattaneo et al., 2006] but the implied outflow rates are enormous. To reproduce the result presented in Figure 3.17, the outflow rate in a halo with $M_h \approx 10^{10.3}\,M_\odot$ must be of order a hundred times higher than the star formation rate. It is difficult at present to reproduce such outflow rates in hydrodynamic simulations. Moreover, observational estimates place the total mass of outflows in normal star-forming galaxies at approximately the same level as the galaxies' final stellar mass [Zahid et al., 2012]. Even in the most extreme observational cases, the "mass-loading factor" (the ratio of mass loss rate due to outflows over the star formation rate) is estimated to be $\lesssim 10$ [Newman et al., 2012]. Therefore, explaining in detail the

mechanisms responsible for the very low η_b values found in low-mass galaxies seems to be a fundamental challenge for models of galaxy formation in a ΛCDM cosmological context.

Recent cosmological hydrodynamic simulations have challenged this statement, by managing to reproduce "realistic" galaxies whose properties satisfy a number of observational constraints. For example, the high-resolution Eris simulation [Guedes et al., 2011] has managed to produce a Milky-Way type object with values of η_* in agreement with those that we see in Fig. 3.17. More recently, McCarthy et al. [2012] has managed to reproduce a population of $\sim 1\,000$ simulated galaxies with low stellar-to-halo mass ratios ($\eta_* \lesssim 0.05$ at $M_h \approx 10^{11.3}\,M_\odot$), in accordance to observations. Notice, however, that this simulation uses a kinetic rather than thermal wind model, in which the initial wind speed is $600\,\mathrm{km/s}$ and the initial mass-loading factor is a factor of four, by construction. Lastly, the work of Guedes et al. [2011] has been extended to lower masses by Brook et al. [2012], who managed to produce a pair of dwarf galaxies ($M_h \approx 10^{10.8}\,M_\odot$), which obey the observed "baryonic Tully-Fisher" relation and are therefore expected to have the correct baryon-to-halo mass fractions.

While these studies indicate that we may be heading toward a solution of the discrepancy between the observed and the expected baryon content of dark matter halos, many questions (e.g. the expected outflow rates and re-accretion timescales) remain unanswered. Therefore, explaining in detail the mechanisms responsible for the very low η_b values found in low-mass galaxies remains an open problem for studies of galaxy formation in a ΛCDM Universe. Furthermore, our measurement of M_b/M_* provides an additional constraint with which cosmological hydrodynamical simulations and semianalytic models will have

to confront.

3.6 Conclusions

We use optical data from the seventh data release of the Sloan Digital Sky Survey (SDSS DR7) and 21cm emission-line data from the Arecibo Legacy Fast ALFA (ALFALFA) survey to measure the "baryonic mass" (defined as $M_b = M_* + 1.4\,M_{HI}$) of galaxies in the local universe, and determine the $z = 0$ baryon mass function (BMF). We use both an HI-selected and an optically-selected sample ($7\,195$ and $22\,587$ galaxies respectively) drawn from the same volume, in order to address the effects of sample selection on the derived distributions. We find that the main difference consists of the optically-selected stellar mass function (SMF) being systematically larger at high-masses than the HI-selected SMF, and find that this difference carries over to the high-mass end of the BMF (see Fig. 3.5 & 3.7).

We combine the obtained mass distributions with the halo mass function in a WMAP3 ΛCDM cosmology, to obtain average values of M_*/M_h and M_b/M_h as a function of halo mass (Fig. 3.16 & 3.17). Our most important result is that low-mass halos seem to have very low galactic baryon fractions compared to the cosmic value ($f_b = \Omega_b/\Omega_m \approx 0.16$), even when their atomic gas content is taken into account; for example, the average baryon fraction of halos with $M_h = 10^{10.3}\,M_\odot$ is just 2% of the cosmic value ($\eta_b \approx 0.02$), and displays a monotonically decreasing trend. This result contrasts with previous indirect measurements of the BMF [Baldry et al., 2008, Evoli et al., 2011], which pointed to an approximately constant value of $\eta_b \approx 0.10$ at the low halo-mass end.

Such very low values of η_b are difficult to reconcile with current models of galaxy formation. Photoionization heating in the early universe suppresses the baryonic content of halos only at $M_h \lesssim 10^{10}\, M_\odot$ [Okamoto et al., 2008], but this mass is more than an order of magnitude smaller than what is required by our result. Therefore, additional feedback mechanisms, such as baryon blowout by supernova explosions, must be present and must be extremely efficient. It is not yet clear whether hydrodynamic simulations or observational results can accommodate such intense galactic outflows in low mass halos. As a result, the observed η_b - M_h relation remains difficult to explain, and may represent a challenge to our understanding of galaxy formation and/or the properties of dark matter.

CHAPTER 4

THE CLUSTERING OF ALFALFA GALAXIES: DEPENDENCE ON HI MASS, RELATIONSHIP TO OPTICAL SAMPLES & CLUES ON HOST HALO PROPERTIES

4.1 Introduction

In the currently accepted hierarchical theory of structure formation, the clustering of galaxies is jointly determined by the large-scale structure of dark matter in the universe, as well as the way in which baryons trace dark matter through the formation of galaxies. As a result, the quantitative study of galaxy clustering through the correlation function, $\xi(r)$, has been instrumental both for constraining cosmological models as well as for furthering our understanding of galaxy formation and evolution.

Given a cosmological model, the clustering of galaxies can be used to constrain the galaxy-halo connection, thus testing and informing models of galaxy formation. A large number of works has been devoted to studying the correlation function of galaxies as a function of their optical properties, such as luminosity, color, morphological and spectral type [e.g. Zehavi et al., 2005, 2011, Li et al., 2012b]. These studies have established with increasing precision a number of fundamental clustering phenomena, such as the trend for stronger clustering with increasing luminosity and the fact that galaxies with blue colors, late-type

*This chapter is an adapted version of the article Papastergis et al. (2013, in prep.), which is in the process of submission.

morphologies and elevated star formation activity cluster significantly less than red, early-type, quiescent galaxies. Moreover, several studies have used the halo occupation distribution (HOD) formalism to make quantitative predictions for the properties of halos hosting a certain class of galaxies. These analyses have suggested that more luminous galaxies inhabit more massive halos on average, and that red galaxies have a higher chance of being hosted by a subhalo compared with blue galaxies. These results are in agreement with theoretical expectations, and are supported by a number of other observational methods (e.g. abundance matching: Guo et al., 2010, Behroozi et al., 2010, Moster et al., 2010, Leauthaud et al., 2012 and galaxy-galaxy weak lensing: Mandelbaum et al., 2006, Dutton et al., 2010, Reyes et al., 2012). These clustering based galaxy occupation models then feed back into cosmological studies, since they provide the necessary link between the measured distribution of galaxies and the distribution of matter that is determined by the cosmological parameters [e.g. Reddick et al., 2012].

Until recently, similarly detailed studies of the clustering characteristics of galaxies selected by their atomic hydrogen content (HI-selected) were not feasible, due to the lack of large-volume blind 21cm surveys. In recent years however, the HI Parkes All Sky Survey (HIPASS; Meyer et al., 2004) and the Arecibo Legacy Fast ALFA[1] (ALFALFA; Giovanelli et al., 2005) survey have provided adequate samples for this purpose. Basilakos et al. [2007] and Meyer et al. [2007] have both analyzed the HIPASS dataset, establishing the fact that HI-selected galaxies are among the most weakly clustered galaxy populations known. In addition, both of these works investigated the dependence of clustering strength on galaxy HI mass (M_{HI}), arriving at different conclusions. More recently, Martin et al. [2012] used $\approx 10\,000$ galaxies from the 40% ALFALFA catalog to measure

the correlation function of gas-rich galaxies. Among their findings were that HI-selected galaxies show a markedly anisotropic clustering pattern (see Fig.4.14 in this chapter), and that they are anti-biased with respect to dark matter on small scales ($\lesssim 5$ Mpc).

In this work, we take advantage of the large HI dataset provided by ALFALFA to make a detailed investigation of the clustering properties of gas-rich galaxies. We furthermore draw samples from the spectroscopic database of the 7th data release of the SDSS [Abazajian et al., 2009] spanning the same volume as the ALFALFA sample, to make comparisons with the clustering properties of optically selected galaxies. The fact that the ALFALFA and SDSS samples are drawn from the same volume allows for a further cross-correlation analysis, measuring the spatial relationship between HI and optical galaxies. Lastly, we select halos from the Bolshoi ΛCDM simulation [Klypin et al., 2011], to investigate what halo properties are associated with weak clustering, giving us evidence on the characteristics of halos hosting gas-rich galaxies.

The chapter is organized as follows: in section 4.2, we present the ALFALFA and SDSS samples used to measure the clustering of HI and optical galaxies, and we describe the methodology for measuring the correlation function. In section 4.3 we present our results concerning the clustering properties of a number of HI-selected and optically selected samples, and discuss the implications. In section 4.4 we present our halo samples selected from the Bolshoi simulation, and study their clustering as a function of their properties (mass, spin, etc.). We conclude in section 4.5, by summarizing the main findings of this work. We forewarn the reader that –unlike most correlation function articles– all distances

[1]The Arecibo L-band Feed Array (ALFA) is a 7-feed receiver operating in the L-band (≈ 1420 MHz), installed at the Arecibo Observatory.

in this work assume a Hubble constant of $H_0 = 70\,h_{70}\ \mathrm{km\,s^{-1}Mpc^{-1}}$. In order to facilitate comparisons with the literature however, the upper x-axis of Figures is expressed in terms of $h \equiv H_0/100\ \mathrm{km\,s^{-1}Mpc^{-1}}$ when appropriate.

4.2 Data & Methods

4.2.1 ALFALFA sample

The ALFALFA survey is a wide-area, blind 21cm emission-line survey performed with the 305m radio telescope at the Arecibo Observatory [Giovanelli et al., 2005]. The survey has recently completed data acquisition, and a source catalog covering ≈40% of the final survey area has been publicly released ("α.40" catalog; Haynes et al., 2011). ALFALFA has greater sensitivity, finer spectral resolution and better centroiding accuracy than previous blind HI surveys of comparable sky coverage (e.g. HIPASS), and α.40 already represents the largest HI-selected galaxy sample to date.

In this chapter, we use a parent sample of 6 123 HI-selected galaxies detected by the ALFALFA survey. In particular, we select galaxies over a contiguous rectangular sky region of $\approx 1\,700$ deg^2 ($135^\circ < \mathrm{RA} < 230^\circ$ and $0^\circ < \mathrm{Dec} < 18^\circ$) and in the redshift range $z \approx 0.0023 - 0.05$ ($\mathrm{v_{CMB}} = 700 - 15\,000\ \mathrm{km\,s^{-1}}$). The parent sample has significant overlap with the publicly available α.40 sample, but has been supplemented by newly processed ALFALFA regions covering the declination ranges $0^\circ < \mathrm{Dec} < 4^\circ$ & $16^\circ < \mathrm{Dec} < 18^\circ$. The sample is restricted to "Code 1" ALFALFA detections, i.e. it is comprised only by confidently detected extragalactic sources ($S/N_{HI} > 6.5$). In addition, parent sample sources have a

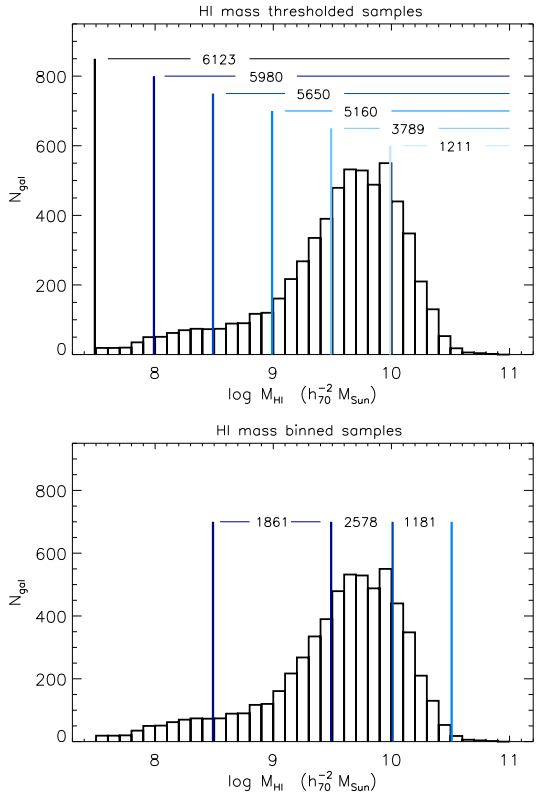

Figure 4.1 The histogram represents the ALFALFA parent sample counts in bins of HI mass. The upper panel is a graphical representation of the HI mass thresholded samples, where each colored vertical line denotes the HI mass threshold of the sample and the total number of galaxies in each sample is quoted. The lower panel is the corresponding plot for the HI mass binned samples. Vertical colored lines denote the HI mass bins limits, while the number of galaxies in each binned sample is also quoted.

combination of observed 21cm flux (S_{HI}) and 21cm lineprofile width (W_{50}) that places them in the region of the $\{S_{HI}, W_{50}\}$–plane where the completeness of the ALFALFA survey is at least 50% [see Sec. 6 and Fig. 12 in Haynes et al., 2011]. Lastly, the sample is limited to linewidths $W_{50} > 18\ \mathrm{km\,s^{-1}}$ and HI masses[2] $M_{HI} > 10^{7.5} M_{\odot}$.

From the parent sample described above we select a number of subsamples, such that their HI mass is above a specified limit (HI mass thresholds) or within a specific range (HI mass bins). Figure 4.1 shows a histogram of M_{HI} for the parent sample, with a graphical representation of the HI mass-thresholded and HI mass-binned subsamples used in this work.

4.2.2 SDSS sample

We select an optical sample of galaxies from the spectroscopic database of the 7th data release of the Sloan Digital Sky Survey (SDSS DR7; Abazajian et al., 2009). This optically selected parent sample is restricted to the same volume as the HI-selected sample used in this work: $135° < \mathrm{RA} < 230°$, $0° < \mathrm{Dec} < 18°$ and $z \approx 0.0023 - 0.05$ ($\mathrm{v_{CMB}} = 700 - 15000\,\mathrm{km\,s^{-1}}$). We only select SDSS galaxies that are spectroscopically classified as galaxies (`SpecClass = 2`) and that have an apparent magnitude in the r-band brighter than 17.6, after correction for Milky Way extinction ($m_r < 17.6$). In addition, we impose a color cut on our spectroscopic sources, $(i - z)_{model} > -0.25$, which excludes a small number of objects; the vast majority of them are cases where star-forming knots and structures in

[2]Atomic hydrogen (HI) masses for ALFALFA galaxies are calculated from their 21cm line flux though the relation $M_{HI} = 2.356\,10^5\ S_{HI}\ d^2$. In this formula M_{HI} is measured in $M_{\odot}$ and the flux S_{HI} in Jy km s^{-1}. The distance d is measured in Mpc, and calculated from the galaxy's recessional velocity in the CMB frame as $d = \mathrm{v}_{CMB}/H_0$.

nearby extended spirals or dwarf irregular galaxies are erroneously classified as separate galaxies by the SDSS pipeline. Lastly, our SDSS parent sample is limited to MW extinction-corrected absolute magnitudes in the r-band brighter than -17 ($M_r < -17$), and is comprised by a total of $18\,516$ galaxies.

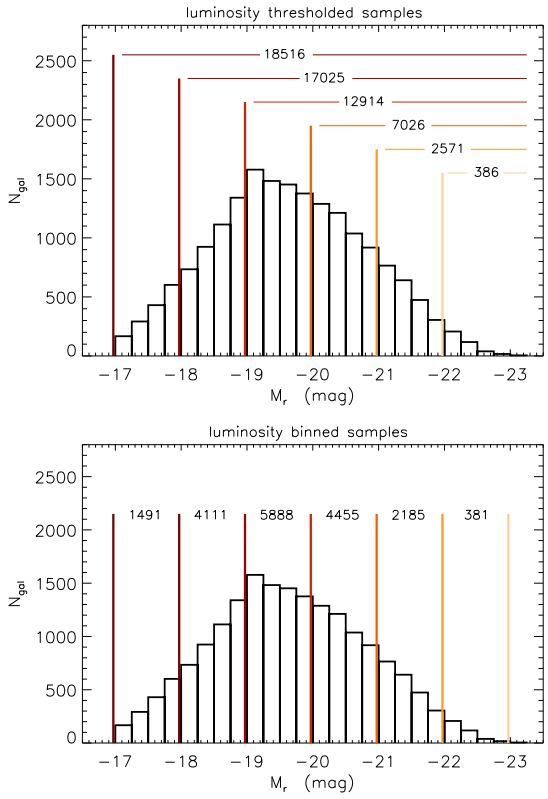

Figure 4.2 Similar to Fig. 4.1, but for the SDSS parent sample. The histogram is the sample count in bins of r-band absolute magnitude, while the upper and lower panels represent graphically the luminosity thresholded and luminosity binned SDSS samples.

From this optical parent sample we create subsamples, selected based on specifying their faintest r-band absolute magnitude (magnitude thresholds) or their range of r-band absolute magnitudes (magnitude bins). Figure 4.2 shows a histogram of M_r for the parent sample, with a graphical representation of the magnitude-thresholded and magnitude-binned subsamples used in this work. Furthermore, we define three color-based subsamples according to the position of galaxies in a color-magnitude diagram (CMD), as shown in Figure 4.3. The

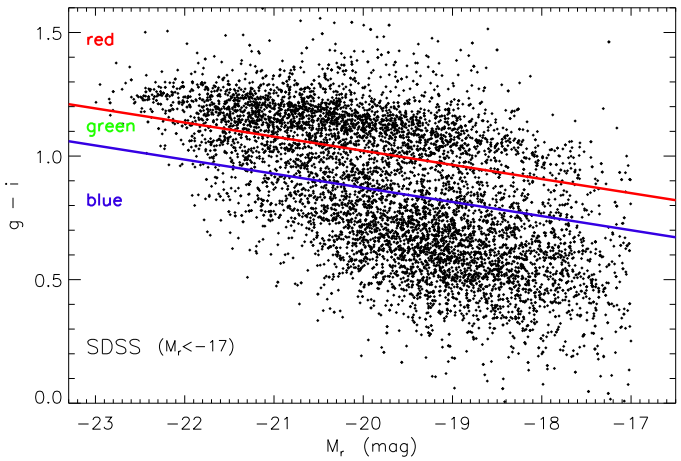

Figure 4.3 Color-magnitude diagram (CMD) of the SDSS parent sample (only a representative subsample is plotted, for visual clarity). The horizontal axis is galactic r-band absolute magnitude, while the vertical axis is the galaxy $g-i$ color. Both quantities are corrected for MW extinction. The solid lines denote the cuts used to select the "red", "green" and "blue" SDSS galaxy samples. The upper red boundary line is given by $g-i = 0.0571(M_r + 24) + 1.25$, while the lower blue boundary line is parallel to the former with a 0.15 mag color offset from it.

"red" subsample is composed by red sequence galaxies, the "blue" subsample by blue cloud galaxies and the green subsample by galaxies with intermediate locations on the CMD, sometimes referred to as "green valley" galaxies.

4.2.3 Sample selection functions & random catalogs

Measuring the clustering of galaxies with certain properties involves comparing the spatial distribution of an observed galactic sample with the spatial distribution of a catalog of random points, which reflect the galactic sample's selection function. The selection function, $\varphi(d)$, describes the fraction of a hypothetical volume-limited sample of galaxies with the desired properties that is included in an observational sample at distance d from the observer. For example, Figure 4.4 shows $\varphi(d)$ for the HI mass-thresholded samples used in this work; sam-

ples restricted to more massive galaxies are complete (i.e. $\varphi = 1$) out to larger distances.

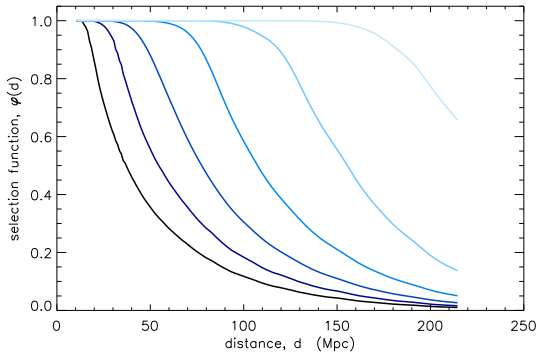

Figure 4.4 Selection functions for the HI mass thresholded samples of Fig. 4.1 (solid lines from bottom to top correspond to samples with $M_{HI} > 10^{7.5}, 10^{8}, 10^{8.5}, 10^{9}, 10^{9.5}, 10^{10}$ $M_{\odot}$). These are the selection functions that are used for constructing the random catalog corresponding to each of the HI mass thresholded samples.

Deriving the selection function for a sample is not straightforward, and necessitates two inputs: (i) the cuts used to define an observational sample and (ii) the intrinsic distribution of the galaxy properties that determine the inclusion of a galaxy in the observational sample. The SDSS sample described in §4.2.2 is mostly flux-limited, because included galaxies satisfy an apparent r-band magnitude cut, $m_r < 17.6$. We thus need to calculate the intrinsic distribution of r-band luminosity for SDSS galaxies, most commonly referred to as the galactic luminosity function (LF). Then $\varphi(d)$ can be calculated in terms of the galaxy luminosity function, $n(M_r)$, as

$$\varphi(d) = \frac{\int_{M_{r,lim}(d)}^{M_{r,min}} n(M_r)\, dM_r}{\int_{M_{r,max}}^{M_{r,min}} n(M_r)\, dM_r} \quad . \tag{4.1}$$

$M_{r,max}$ and $M_{r,min}$ are the faint and bright absolute magnitude limits defining a specific subsample, while $M_{r,lim}(d)$ is the faintest absolute magnitude that a

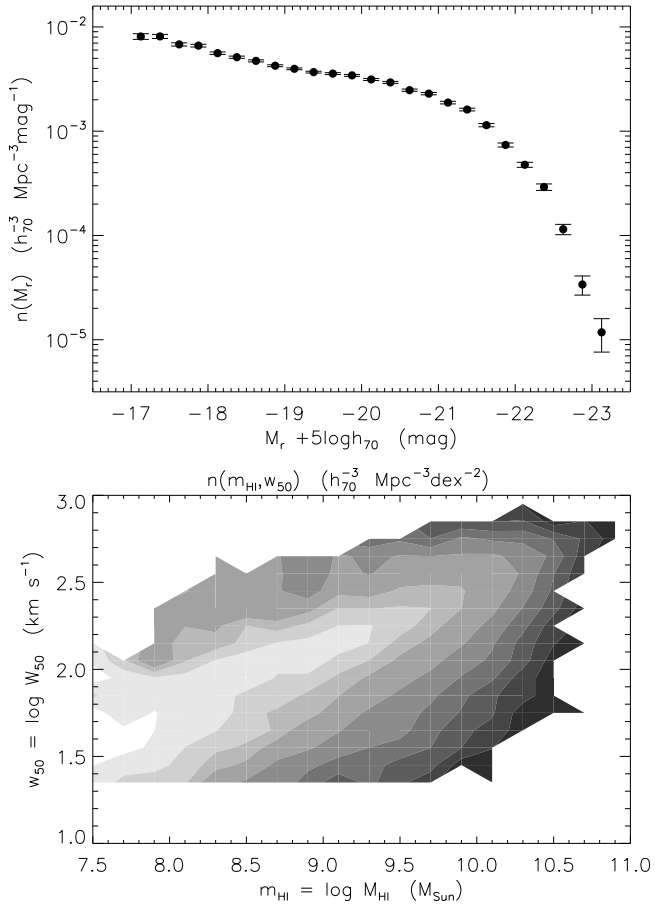

Figure 4.5 *upper panel:* The SDSS r-band luminosity function, used to calculate the selection function for the SDSS subsamples used in this work. The error bars denote only the Poisson error due to the number of sample galaxies in each M_r bin. *lower panel:* The ALFALFA mass-width function, used to calculate the selection function for the ALFALFA subsamples used in this work. The contours are set at $n(m_{HI}, w_{50}) = 10^{-6}, 10^{-5.5}, \ldots, 10^{-2.5}, 10^{-1.75}, \ldots, 10^{-1.25}$ $\mathrm{Mpc}^{-3}\mathrm{dex}^{-2}$, from darker to lighter tones.

galaxy at distance d can have and still have an apparent magnitude brighter than $m_r = 17.6$. The luminosity function, $n(M_r)$, is the volume-limited number density of galaxies within a bin of magnitude centered on M_r, and has units of $\mathrm{Mpc}^{-3}\,\mathrm{mag}^{-1}$. In informal terms, the denominator in Eqn. 4.1 represents the volume-limited number density of a specific subsample, while the numerator represents the number density of galaxies in the subsample that are detectable at distance d.

On the other hand the ALFALFA sample described in §4.2.1 is not a purely flux-limited sample, but it is mostly defined through a flux-width–dependent cut [Eqns. 4 & 5 in Sec. 6 of Haynes et al., 2011]. We therefore need to know the intrinsic two-dimensional distribution of HI mass and linewidth, $n(m_{HI}, w_{50})$, of ALFALFA galaxies. The mass-width-function is customarily expressed in logarithmic intervals of mass and width, so $m_{HI} = \log(M_{HI}/M_\odot)$ and $w_{50} = \log(W_{50}/\mathrm{km\,s^{-1}})$. The selection function for any ALFALFA subsample is then given by the expression:

$$\varphi(d) = \frac{\int_{w_{50,min}}^{w_{50,max}} \int_{m_{HI,lim}(d,w_{50})}^{m_{HI,max}} n(m_{HI}, w_{50})\, dm_{HI}\, dw_{50}}{\int_{w_{50,min}}^{w_{50,max}} \int_{m_{HI,min}}^{m_{HI,max}} n(m_{HI}, w_{50})\, dm_{HI}\, dw_{50}} \quad . \tag{4.2}$$

Again, $m_{HI,min}$, $m_{HI,max}$, $w_{50,min}$ and $w_{50,max}$ are the HI mass and linewidth limits defining a specific ALFALFA subsample, while $m_{HI,lim}(d, w_{50})$ is the minimum HI mass detectable at distance d for a source of linewidth w_{50}, as dictated by the ALFALFA 50% completeness limit. $n(m_{HI}, w_{50})$ is again the number density of galaxies within a logarithmic bin of HI mass centered on m_{HI} and a logarithmic bin of width centered on w_{50}, and has units of $\mathrm{Mpc}^{-3}\,\mathrm{dex}^{-2}$.

The r-band luminosity function for SDSS galaxies and the mass-width-

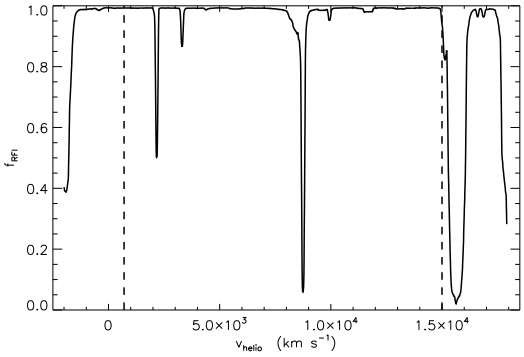

Figure 4.6 The black solid line represents the fraction of the nominal surveyed volume that is available for ALFALFA source extraction in the presence of radio frequency interference (RFI), as a function of heliocentric velocity (roughly equivalent to antenna rest frequency). The largest dip at $v_\odot \approx 16\,000\ \mathrm{km\,s^{-1}}$ is due to the San Juan airport radar, while the second largest dip at $v_\odot \approx 8\,800\ \mathrm{km\,s^{-1}}$ is one of the radar's harmonics. The vertical dashed lines are the approximate redshift limits of the ALFALFA parent sample used in this chapter; the high redshift limit has been specifically chosen so as to avoid the strongest RFI peak.

function for ALFALFA galaxies used in Eqns. 4.1 & 4.2 are shown in Figure 4.5. They are calculated by applying appropriate volume correction factors to the sample histograms of M_r and $\{m_{HI}, w_{50}\}$, respectively. The volume weights are calculated with the maximum-likelihood, non-parametric "$1/V_{eff}$" method [Zwaan et al., 2005], on a galaxy-by-galaxy basis. For a more detailed description of the method see the Appendix.

Once a data sample selection function is known, it is straightforward to construct a random catalog suitable for the calculation of the sample correlation function. Initially, random points are created within the subsample volume with a constant expected number density throughout, $\langle dN_{rand}/dV \rangle = \mathrm{const.}$ This translates into random points being uniformly distributed in RA, $\sin(\mathrm{Dec})$ and d^3. Subsequently, each random point is kept with probability $\varphi(d)$ (where d is the random point's distance from the observer), in order to reproduce the subsample's selection function. In the case of an ALFALFA sample an additional step is

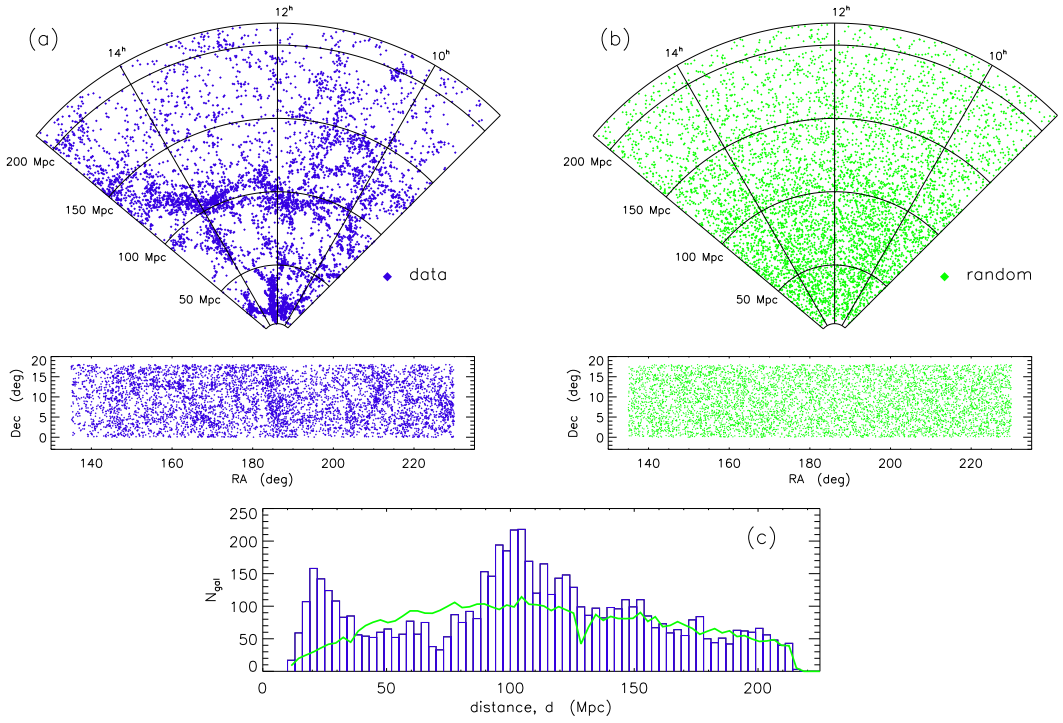

Figure 4.7 *panel (a):* Coneplot (polar plot of distance & RA) and sky position (rectangular plot of RA & Dec) of the ALFALFA parent sample (blue diamonds). *panel (b):* Same as panel (a), but for the corresponding catalog of random points (green diamonds). *panel (c):* Distance distribution of the ALFALFA parent sample (blue histogram) and its corresponding random catalog (green line). Note the markedly non-uniform distribution of ALFLAFA galaxies, which is evident in all three plots.

necessary: accounting for the effects of radio frequency interference (RFI). RFI disrupts ALFALFA's performance in the frequency bands where it occurs, resulting in galaxies with certain heliocentric velocities having a lower chance of being detected. Figure 4.6 shows the fractional ALFALFA volume lost to RFI as a function of heliocentric velocity. In order to reproduce the effects of RFI on the spatial distribution of ALFALFA samples, points in the random catalog are kept with a probability $f_{RFI}(\mathrm{v}_\odot)$ where $\mathrm{v}_\odot$ is the heliocentric velocity of the random point[3].

Figure 4.7 compares the distribution of data and random points for the $M_{HI} = 10^{9.5} - 10^{10}\,M_\odot$ ALFALFA sample. Panel (a) displays the coneplot (i.e. a projection of RA and d in polar coordinates) and the sky distribution of the data sample, while panel (b) displays the same distributions for the corresponding random catalog. Panel (c) compares the distance histograms of the two samples. The non-uniform distribution of the data set and the large-scale structure in the survey volume are readily visible in all three panels.

4.2.4 Clustering measures

The galaxy correlation function at a given length scale, $\xi(r)$, is defined as the excess probability of finding a pair of galaxies separated by distance r compared to the case of a randomly distributed set of points. It then follows that a positive value of $\xi(r)$ means that the sample under consideration tends to cluster on length scales r, while a negative value means that it avoids clus-

[3]The heliocentric velocity of a random point is calculated by first considering their velocity relative to the cosmic microwave background (CMB), $\mathrm{v}_{CMB} = H_0\,d$, and then converting the velocity from the CMB to the heliocentric frame according to the random point's position in the sky.

tering on this scale (a randomly distributed sample would have $\xi(r) = 0$ for all r). In formal terms, the correlation function is defined through the relation $d^2N_{pair} = \bar{n}^2_{gal}\ \xi(r)\, P_{rand}(r)\ dV_1\, dV_2$. Here dV_1 and dV_2 are two volume elements separated by distance r, d^2N_{pair} is the number of galaxy pairs within those volume elements, $\bar{n}_{gal}$ is the average galaxy number density and $P_{rand}(r)$ is the probability that two random points are separated by distance r. In practice, when a galactic sample and its corresponding random catalog are available, the correlation function is calculated in terms of the number of data-data, random-random and data-random pairs whose separation falls in the bin $r \pm \Delta r/2$. These pair counts are denoted by $P_{DD}(r), P_{RR}(r)$ & $P_{DR}(r)$, respectively. If the data sample contains N_D objects and the random sample N_R objects, we can compute the normalized counts

$$\begin{aligned} DD(r) &= P_{DD}(r) \,/\, (N_D(N_D-1)/2) \\ RR(r) &= P_{RR}(r) \,/\, (N_R(N_R-1)/2) \\ DR(r) &= P_{DR}(r) \,/\, (N_D N_R)\ , \end{aligned} \tag{4.3}$$

where in all three cases the denominator represents the total number of available pairs.

The most intuitive estimator for the correlation function is then $\hat{\xi}(r) = DD(r)/RR(r) - 1$, which just computes the ratio of the fraction of data-data pairs and random-random pairs separated by distance r and compares it with unity. However, Landy and Szalay [1993] have shown that an alternative estimator, $\hat{\xi}_{LS}(r) = (DD(r) - 2DR(r) + RR(r)) \,/\, RR(r)$, has better statistical performance; for volume-limited, weakly clustered samples equipped with large random cat-

alogs ($N_R \gg N_D$) the $\hat{\xi}_{LS}$ estimator is unbiased and its variance is determined just by the counting noise associated with the number of data-data pairs. In this chapter we adopt throughout the LS estimator, dropping from now on the excess notation: $\hat{\xi}_{LS}(r) \rightarrow \xi(r)$.

Despite the fact that the "real space" correlation function, $\xi(r)$, is the fundamental quantity related to galaxy clustering, physical separation is not generally available for extragalactic objects. The measurable quantities in a spectroscopic galaxy survey are position on the sky $(\mathrm{RA}, \mathrm{Dec})$ and recessional velocity ($\mathrm{v}_{CMB} = c\,z$). As a result we consider in this chapter the "redshift space" separation between two objects s, given by

$$s = \sqrt{(v_1^2 + v_2^2 - 2v_1 v_2 \cos\theta)}/H_0 \quad , \tag{4.4}$$

where v_1, v_2 are the recessional velocities of galaxies 1 and 2 respectively in $\mathrm{km\,s^{-1}}$, θ is the angle between them on the sky, and H_0 is the Hubble constant (recall that in this chapter $H_0 = 70\ \mathrm{km\,s^{-1}}$ Mpc^{-1}). In addition we can consider separately the components of the separation along the line of sight (π) and on the plane of the sky (σ) defined as:

$$\pi = |v_1 - v_2| \,/\, H_0 \quad \text{and} \tag{4.5}$$

$$\sigma = \sqrt{s^2 - \pi^2}\,. \tag{4.6}$$

We can therefore calculate the redshift space correlation function $\xi(s)$ by counting the number of pairs whose separation is within $s \pm \Delta s/2$. Similarly, we can calculate the two-dimensional correlation function $\xi(\sigma, \pi)$ by counting pairs sep-

arated by $\sigma \pm \Delta\sigma/2$ in the tangential plane and $\pi \pm \Delta\pi/2$ along the line of sight. Note that in the absence of galaxy peculiar velocities (i) $\xi(s)$ would coincide with $\xi(r)$ and (ii) $\xi(\sigma, \pi)$ would contain no additional information compared to $\xi(s)$, since galaxy clustering is expected to be intrinsically isotropic. However, due to "redshift-space distortions" $\xi(\sigma, \pi)$ has a characteristic non-isotropic shape (see Fig. 4.14), and contains non-trivial information regarding cosmic properties [see e.g. Reid et al., 2012, Contreras et al., 2013].

Lastly, we can measure the "projected correlation function", which is denoted by $\Xi(\sigma) \,/\, \sigma$ and is defined as[4]

$$\Xi(\sigma) \,/\, \sigma = \frac{2}{\sigma} \int_{\pi=0}^{\pi=\pi_{max}} \xi(\sigma, \pi)\, d\pi \quad , \tag{4.7}$$

where $\pi_{max} = 45\ h_{70}^{-1}$ Mpc is used in this chapter. $\Xi(\sigma) \,/\, \sigma$ is a correlation measure that is integrated over the line-of-sight direction. As a result, it is not affected by redshift space distortions and therefore is the most closely related to $\xi(r)$. In fact, if the real space correlation function follows a power-law form, parametrized as $\xi(r) = (r \,/\, r_0)^{-\gamma}$, then

$$\Xi(\sigma) \,/\, \sigma = \frac{r_0^{\gamma} \Gamma(1/2) \Gamma((\gamma-1)/2)}{\Gamma(\gamma/2)} \ \ \sigma^{-\gamma} \quad . \tag{4.8}$$

In other words, a power-law projected correlation function has the same exponent γ as the real space correlation function, while at the same time its normalization can be used to determine the clustering scale-length parameter, r_0.

[4]There exists a different notation and definition for the projected correlation function that is more widely used in the literature, $w_p(r_p)$. In this case r_p is the separation on the plane of the sky (same as σ) and $w_p = 2 \int_{\pi=0}^{\pi=\pi_{max}} \xi(r_p, \pi)\, d\pi$. Therefore, the two definitions differ only by a factor of $r_p \equiv \sigma$. In this chapter, we opt for $\Xi(\sigma) \,/\, \sigma$ because (i) it is a unitless quantity and

4.2.5 Pair-weighting

In order to increase the effective volume probed by the ALFALFA and SDSS samples we weight each pair roughly inversely to the product of the individual selection function values for the two constituent objects. This weighting aims at taking into account the large number of pairs that remain undetected at large distances. More specifically, each data-data, random-random and data-random pair is counted towards P_{DD}, P_{RR} and P_{DR} with a weight w_{ij} given by

$$w_{ij} = w_i \times w_j \ , \quad \text{where} \tag{4.9}$$

$$w_i = \frac{1}{1 + 4\pi\bar{n}J_3 \ \varphi(d_i)} \ . \tag{4.10}$$

In the expression above, w_i and w_j are the weights of object i and j respectively, while $\varphi(d_i)$ is the selection function at the distance of object i. $\bar{n}$ is the average volume-limited number density of the sample, while J_3 is a short hand notation for $J_3(s = 30Mpc) = \int_{s=0}^{s=30\mathrm{Mpc}} s^2\,\xi(s)\,ds$. Results are not sensitive to the exact value of J_3, so a value of $J_3 = 2\,962\ \mathrm{Mpc}^3$ is used here, corresponding to a fiducial $\xi_{fid}(s) = (s/5\,\mathrm{Mpc})^{-1.5}$. In essence, the weight in Eqn. 4.10 reduces to $w_{ij} \propto \frac{1}{\varphi_i} \times \frac{1}{\varphi_j}$ when the selection function is relatively large ($\varphi(d) \gg 1/(4\pi\bar{n}J_3)$), while $w_{ij} \approx 1$ when the selection function is small ($\varphi(d) \ll 1/(4\pi\bar{n}J_3)$).

In the case of SDSS data-data pair counts, an additional weight is applied to correct for SDSS "fiber collisions". SDSS spectroscopic fibers cannot generally be placed closer than $55''$ from one another, lowering the counts of pairs at small on-sky separations. As a result, SDSS data-data pair weights are given by

therefore independent of H_0 and (ii) has the same logarithmic slope as the real space correlation function, $\xi(r)$.

$$w_{ij} = w_i \times w_j \times w_{fc}(\theta_{ij}) \ . \tag{4.11}$$

w_i and w_j are defined as per Eqn. 4.10, while $w_{fc}(\theta_{ij})$ is the fiber collision correction that depends only on the angular separation between the two galaxies. The analytic form of $w_{fc}(\theta)$ used in this chapter (Cheng Li, private communication) is the same as the one described in Li et al. [2006a] and tested with mock catalogs in Li et al. [2006b]. Note however that the correlation functions presented in this chapter are limited to scales $\gtrsim 200$ kpc, and so the impact of fiber collisions is small.

4.2.6 Error estimation

We calculate errors on our clustering measurements by bootstrapping our data sample. If a data sample has N_D elements, bootstrap resampling involves forming sample realizations by randomly extracting N_D elements from the original data set, with replacement. If $k = 1, \ldots, K$ sample realizations are produced in this way, we can calculate statistical properties of the measured correlations such as the average, variance and covariance matrix:

$$\langle \xi_i \rangle = \frac{1}{K} \sum_{k=1}^{K} \xi_i^{(k)} \tag{4.12}$$

$$\sigma_{\xi_i}^2 = \frac{1}{K-1} \sum_{k=1}^{K} (\xi_i^{(k)} - \langle \xi_i \rangle)^2 \ \text{ and} \tag{4.13}$$

$$\mathrm{Cov}(\xi_i, \xi_j) = \frac{1}{K-1} \sum_{k=1}^{K} (\xi_i^{(k)} - \langle \xi_i \rangle)(\xi_j^{(k)} - \langle \xi_j \rangle) \ . \tag{4.14}$$

Here, ξ_i denotes generically the value of some correlation measure in separation bin i, while the superscript $^{(k)}$ denotes the $k^{\rm th}$ bootstrap sample realization. Note that, as is usual practice, we use in this chapter random catalogs with more objects than our data samples ($N_R > N_D$). This is done in order to ensure that the contribution of the counting noise of random-random pairs to the overall error budget is subdominant to the error from data-data and data-random pairs. Specifically, each random catalog is 10 times the length of the corresponding data catalog (up to a maximum of $100\,000$), and 25 bootstrap realizations are used to estimate the mean, variance and covariance. Finally, to ease the computational workload, the normalized random-random counts are only computed once, while the data-data and data-random pairs are computed for each realization.

4.3 Results

4.3.1 Dependence of clustering on HI mass

Figure 4.8 shows the measured projected correlation function, $\Xi(\sigma)\,/\,\sigma$, for the HI mass thresholded samples shown in the upper panel of Figure 4.1, and described in §4.2.1. A few preliminary comments are worth making: Firstly, the projected correlation function for all samples is well approximated by a power-law, up to a length scale of $\sigma \approx 15\ \mathrm{h}_{70}^{-1}\,\mathrm{Mpc}$. Secondly, the correlation functions seem to deviate from the simple power-law form at separations larger than this characteristic value. This behavior has been noted in multiple literature studies and seems to hold for both optically selected and HI-selected samples [e.g.

Li et al., 2012b, Norberg et al., 2002, Zehavi et al., 2011 for optical samples and Martin et al., 2012, Basilakos et al., 2007 for HI samples].

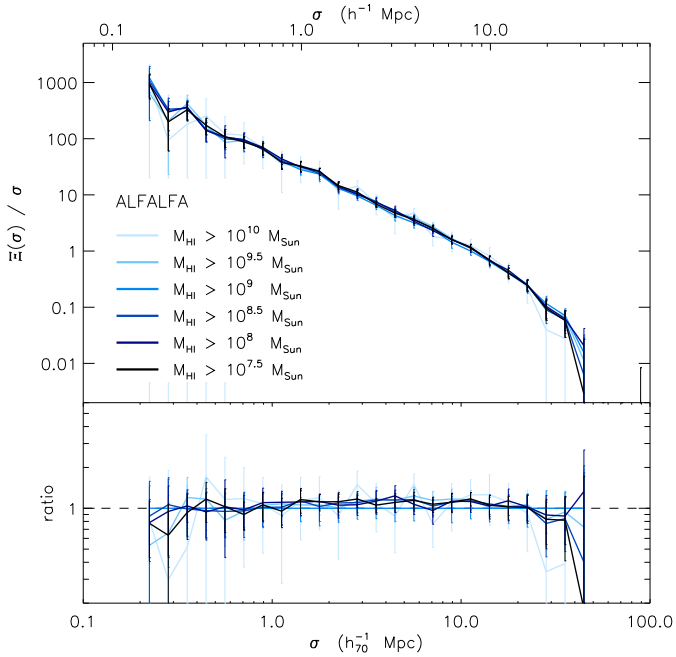

Figure 4.8 *upper panel:* The projected correlation function, $\Xi(\sigma)\,/\,\sigma$, for the HI mass thresholded samples shown in the upper panel of Figure 4.1 and described in §4.2.1. Darker shades of blue represent samples with a lower HI mass threshold, as depicted in the Figure legend. Note that some error bars for the $M_{HI} > 10^{10}\ M_{\odot}$ sample extend below the legend. *lower panel:* The same projected correlation functions as above, normalized to the correlation function of the $M_{HI} > 10^{9}\ M_{\odot}$ sample. The unity line is also plotted for reference.

Most importantly however, the correlation functions of the HI mass thresholded samples show no significant differences among one another, within the errors of the present analysis. Note that Figure 4.8 shows no evidence for enhanced clustering for the samples with the highest HI masses; this is in stark contrast to the strong clustering displayed by galaxies with high stellar mass or optical luminosity [e.g. Zehavi et al., 2011, Beutler et al., 2013, to name a few, see also Fig.4.11 in this work]. Figure 4.8 is not ideal however for assessing the clustering properties of low HI mass galaxies; as the upper panel of Fig-

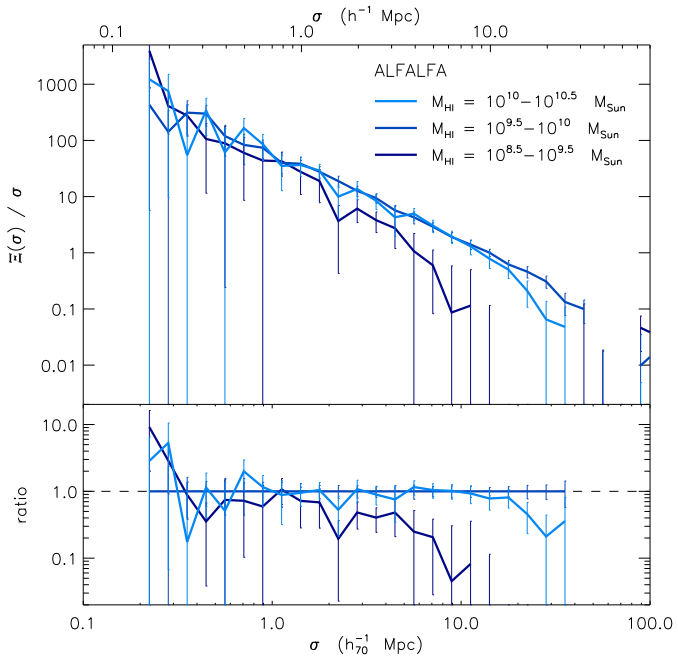

Figure 4.9 *upper panel:* The projected correlation function, $\Xi(\sigma)\,/\,\sigma$, for the non-overlapping HI mass binned samples shown in the lower panel of Figure 4.1 and described in §4.2.1. Darker shades of blue represent samples with a lower range of HI masses, as depicted in the Figure legend. *lower panel:* The same projected correlation functions as above, normalized to the correlation function of the $M_{HI} = 10^{9.5} - 10^{10}\ M_{\odot}$ sample. The unity line is also plotted for reference. Note that the range of the y-axis in the lower panel of this Figure is much larger than in Figure 4.8.

ure 4.1 shows, even the samples with the lowest HI mass thresholds (as low as $M_{HI} = 10^{7.5}\ M_{\odot}$) are still dominated by fairly HI massive galaxies. We therefore display in Figure 4.9 the $\Xi(\sigma)\,/\,\sigma$ measurements for the three non-overlapping HI mass binned samples shown in the lower panel of Figure 4.1. Galaxies with intermediate and high masses ($M_{HI} = 10^{9.5} - 10^{10}\ M_{\odot}$ and $M_{HI} = 10^{10} - 10^{10.5}\ M_{\odot}$ bins) show no significant differences in their clustering properties in Figure 4.9, in accordance with the results in Figure 4.8. Interestingly enough though, galaxies with low HI mass ($M_{HI} = 10^{8.5} - 10^{9.5}\ M_{\odot}$ bin) seem to be much more weakly clustered than their more massive counterparts.

The HI mass dependence of the clustering properties of HI-selected galaxies has remained a controversial issue in the literature. For instance, Basilakos et al. [2007] and Meyer et al. [2007] have both analyzed datasets from the HIPASS survey, but came to different conclusions regarding the issue. On one hand, Basilakos et al. [2007] found that HIPASS galaxies with $M_{HI} < 10^{9.4}\ M_{\odot}$ have a significantly lower clustering amplitude than galaxies with HI masses larger than this value (see their Fig. 5). On the other hand, Meyer et al. [2007] dissected the HIPASS sample at a similar HI mass, $M_{HI} = 10^{9.25}\ M_{\odot}$, but found no convincing differences in the correlation function of the low-mass and high-mass subsamples (see their Fig. 12). At face value, the ALFALFA measurement shown in our Figure 4.9 seems to lend support to the Basilakos et al. [2007] claim. One complication arises however due to the fact that the volume probed by the $M_{HI} = 10^{8.5} - 10^{9.5}\ M_{\odot}$ sample is ≈ 6 times smaller than the volume probed by the $M_{HI} = 10^{9.5} - 10^{10}\ M_{\odot}$ sample. As a result the observed discrepancy could be caused by finite volume effects. We therefore re-calculate the projected correlation function of the $M_{HI} = 10^{9.5} - 10^{10}\ M_{\odot}$ sample, but restricting it to the smaller volume available to the $M_{HI} = 10^{8.5} - 10^{9.5}\ M_{\odot}$ sam-

ple. Figure 4.10 shows the result: Even though the correlation functions of the two samples are very different from one another when both are calculated over their full volumes (dark blue and blue solid lines), they show no significant differences when the two samples are restricted to a common volume (dark blue solid line and blue dash-dotted line).

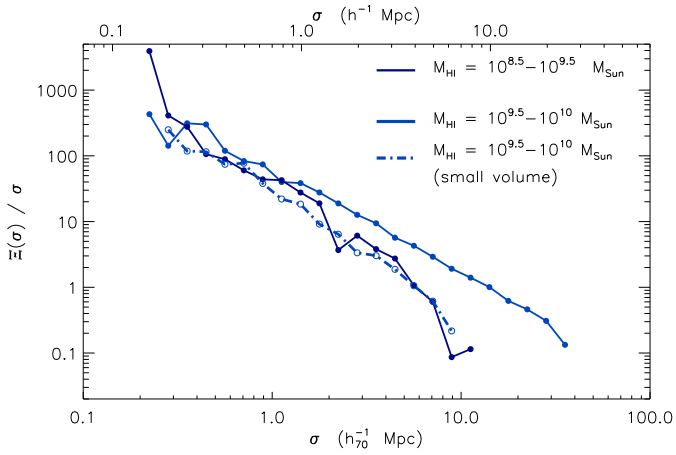

Figure 4.10 The solid lines are the projected correlation functions for the $M_{HI} = 10^{8.5} - 10^{9.5}\ M_\odot$ and $M_{HI} = 10^{9.5} - 10^{10}\ M_\odot$ samples (darker and lighter shade, respectively). The lighter shade dash-dotted line shows again the correlation function of the $M_{HI} = 10^{9.5} - 10^{10}\ M_\odot$ sample, but this time restricted to the ≈ 6 times smaller volume occupied by the galaxies in the $M_{HI} = 10^{8.5} - 10^{9.5}\ M_\odot$ sample. Error bars are omitted for clarity. This Figure shows that the difference in clustering between the two samples is probably due entirely to finite volume effects.

Overall, we find no conclusive evidence for a dependence of the clustering properties of HI-selected galaxies on their HI mass, over the mass range $M_{HI} \approx 10^{8.5} - 10^{10.5}\ M_\odot$. Despite the fact that Figure 4.9 displays a weak correlation function for low HI mass galaxies ($M_{HI} = 10^{8.5} - 10^{9.5}\ M_\odot$), Figure 4.10 suggests that this behavior could be entirely due to finite volume effects. An extension of this work to both higher and lower masses will necessitate the next generation of HI surveys, such as the planned WALLABY survey with the ASKAP array [Koribalski, 2012] and the HI surveys to be performed with the APERTIF instrument on the WSRT interferometer, which will probe a much

larger volume than ALFALFA and are expected to detect ≈10x more sources. These surveys will also provide more accurate measurements of the correlation function over the HI mass range probed in this work, potentially uncovering trends that cannot be detected within current precision.

4.3.2 Bias relative to optical galaxies

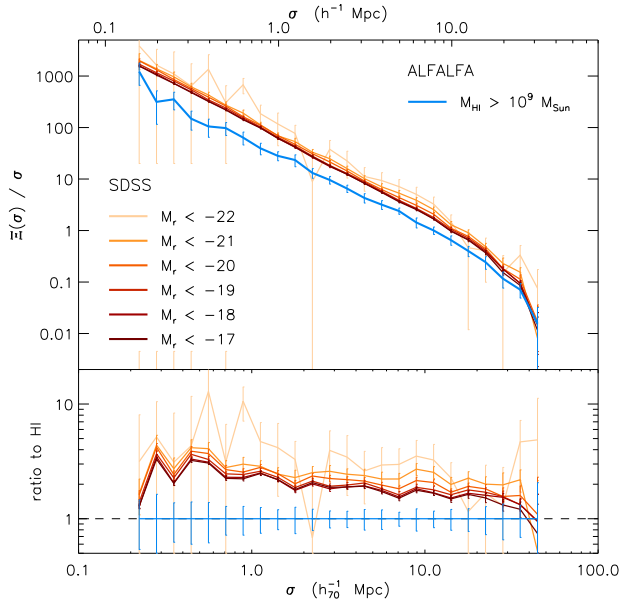

Figure 4.11 *upper panel:* The projected correlation function, $\Xi(\sigma)\,/\,\sigma$, for the luminosity thresholded SDSS samples (shown in the upper panel of Figure 4.2 and described in §4.2.2), compared to the correlation function of the $M_{HI} > 10^9\ M_\odot$ ALFALFA sample. Darker shades of red represent samples with a lower optical luminosity threshold, as depicted in the Figure legend. Note that some of the error bars of the $M_r < -22$ sample extend below the legend. *lower panel:* The same projected correlation functions as above, normalized to the correlation function of the ALFALFA sample. The unity line is also plotted for reference.

Several literature studies have found that HI-rich galaxies are among the most weakly clustered galactic populations known [e.g. Basilakos et al., 2007, Meyer et al., 2007, Martin et al., 2012]. This fact can be clearly seen in Figure 4.11, which compares the projected correlation function of one represen-

tative ALFALFA sample ($M_{HI} > 10^9\ M_\odot$) with the correlation functions of the luminosity thresholded SDSS samples (as depicted in the upper panel of Figure 4.2 and described in §4.2.2). The Figure shows that a "typical" HI-selected sample is significantly less clustered than a "typical" optically selected sample, regardless of the optical sample's limiting luminosity.

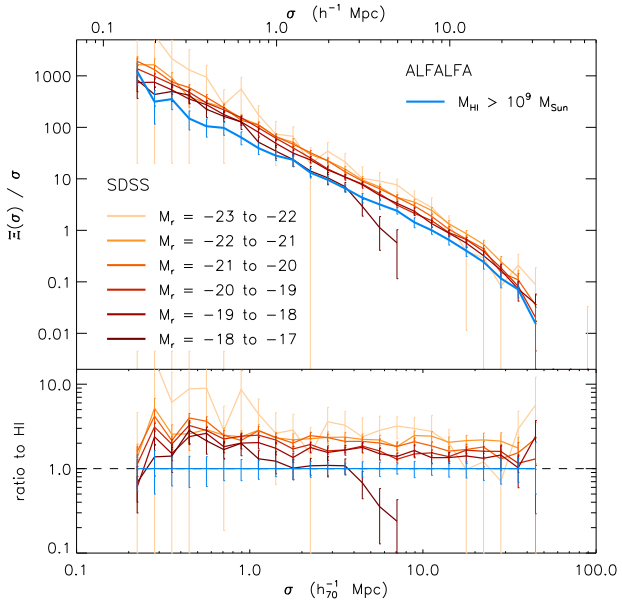

Figure 4.12 The projected correlation function, $\Xi(\sigma)\,/\,\sigma$, for the luminosity binned SDSS samples (shown in the lower panel of Figure 4.2 and described in §4.2.2), compared to the correlation function of the $M_{HI} > 10^9\ M_\odot$ ALFALFA sample. Darker shades of red represent samples with a lower range of optical luminosities, as depicted in the Figure legend. Note that some of the error bars for the M_r = -23 – -22 sample extend below the legend. *lower panel:* The same projected correlation functions as above, normalized to the correlation function of the ALFALFA sample. The unity line is also plotted for reference.

Furthermore, we compare the correlation function of the same ALFALFA sample to the correlation function of the SDSS luminosity-binned samples (as depicted in the lower panel of Figure 4.2). The result is shown in Figure 4.12, on which we note the following points: Firstly, the optical samples display a clear trend of stronger clustering with increasing luminosity, unlike HI mass-binned samples (Fig. 4.9). The dependence of clustering on luminosity has been exten-

sively studied in the literature, and the observed trend has been interpreted as a tendency of more luminous galaxies to inhabit more massive DM halos [see e.g. Zehavi et al., 2011, Beutler et al., 2013 for two recent examples]. Secondly, and most importantly, the correlation function of the HI-selected sample is lower in amplitude than the correlation function of even relatively faint optical galaxies (at least as faint as $M_r \approx -18$). In addition, the optically selected samples seem to display a steeper correlation function regardless of luminosity.[5]

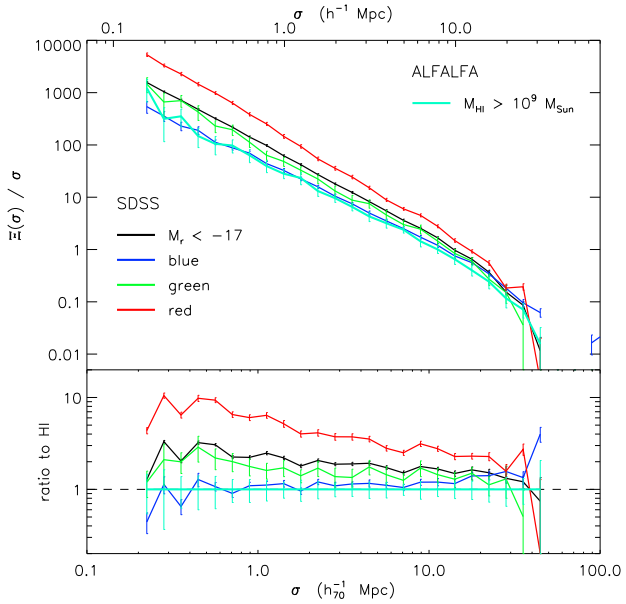

Figure 4.13 *upper panel:*The projected correlation function, $\Xi(\sigma)\,/\,\sigma$, for the "blue", "green" & "red" SDSS samples (shown in Figure 4.3 and described in §4.2.2), compared to the correlation function of the $M_{HI} > 10^9\ M_\odot$ ALFALFA sample. *lower panel:* The same projected correlation functions as above, normalized to the correlation function of the ALFALFA sample. The unity line is also plotted for reference.

Figure 4.13, on the other hand, compares the clustering of the same $M_{HI} > 10^9\ M_\odot$ ALFALFA sample to the clustering of three optical subsamples split by color (see Fig. 4.3 and §4.2.2). By combining the information in Figs. 4.11, 4.12

[5]Note that the $M_r = -18$ to -17 sample has not been taken into account when making the statements above. The correlation function of this sample is probably affected significantly by finite-volume effects, similarly to the case of the $M_{HI} = 10^{8.5} - 10^{9.5}\ M_\odot$ ALFALFA sample (see Fig. 4.10).

& 4.13, we arrive at the following conclusions:

1. HI-selected galaxies cluster less than optically selected galaxies, when no color cuts are applied to the latter sample. This statement is valid even for relatively faint galaxies (at least as faint as $M_r \approx -18$). In addition, optically selected samples of all luminosities display slightly steeper correlation functions compared to HI-selected samples.
2. The correlation function of HI-selected galaxies is practically indistinguishable from the correlation function of optical galaxies with blue colors. The relative bias[6] of the two samples is $b_{rel} \approx 1$, over almost the whole range of separations probed.
3. Red galaxies show much stronger clustering than HI-selected galaxies, with the relative bias reaching values $b_{rel} > 3$ at small separations ($\sigma \lesssim 1$ Mpc). Moreover, the projected correlation function of red optical galaxies is significantly steeper than that of HI-selected galaxies.

Points 1-3 above hold also for the full two-dimensional correlation function, $\xi(\sigma, \pi)$: Figure 4.14 shows $\xi(\sigma, \pi)$ for the parent ALFALFA sample ($M_{HI} > 10^{7.5}\ M_{\odot}$), which can be compared with $\xi(\sigma, \pi)$ for the blue and red SDSS subsamples (Figures 4.15 & 4.16, respectively). Note that common contour levels are used in Figs. 4.14 - 4.16. The two-dimensional correlation functions for the ALFALFA and blue SDSS galaxies are very similar in amplitude and shape. In particular, both samples display a characteristic "flattening" of $\xi(\sigma, \pi)$ along the π-axis on intermediate scales ($\pi \gtrsim 10$ Mpc), as well as a weak "finger of god"

[6]The relative bias between two samples s_1 & s_2 is defined as the square root of the ratio of their real space correlation functions, in other words $b^2_{rel}(r) \equiv \xi_{s_1}(r)/\xi_{s_2}(r)$. Bias values quoted in this article are calculated by fitting the projected correlation function with Eqn. 4.8, under the assumption of a power-law form for the real space correlation function, $\xi(r)$.

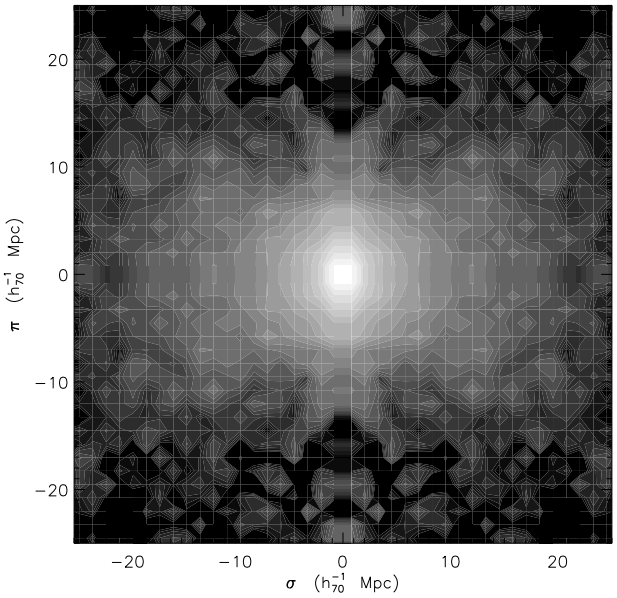

Figure 4.14 The two-dimensional correlation function, $\xi(\sigma, \pi)$, of the ALFALFA parent sample ($M_{HI} > 10^{7.5}\ M_{\odot}$). Note that $\xi(\sigma, \pi)$ is calculated in *linear* bins of separation, with $\sigma_{min} = \pi_{min} = 0.15\ h_{70}^{-1}\,\mathrm{Mpc}$ and bin size $\Delta\sigma = \Delta\pi = 1.25\ h_{70}^{-1}\,\mathrm{Mpc}$. The contours are logarithmically spaced, starting at a value of 0.05 and increasing by a factor of 2 every three contours up to a factor of 6.3. Note also that the full information of $\xi(\sigma, \pi)$ is contatined in one quadrant of the plot only; the other quadrants are just mirrored copies.

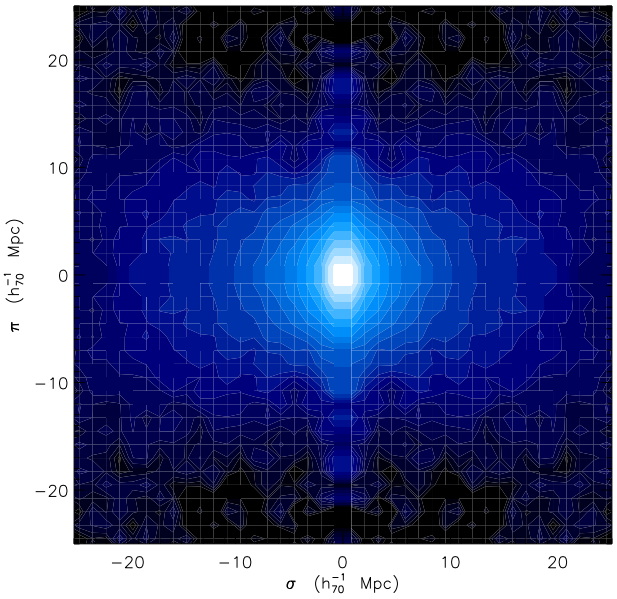

Figure 4.15 The two-dimensional correlation function, $\xi(\sigma, \pi)$, of the SDSS "blue" sample (see Fig. 4.3). The separation bins and contour levels are the same as for Fig. 4.14.

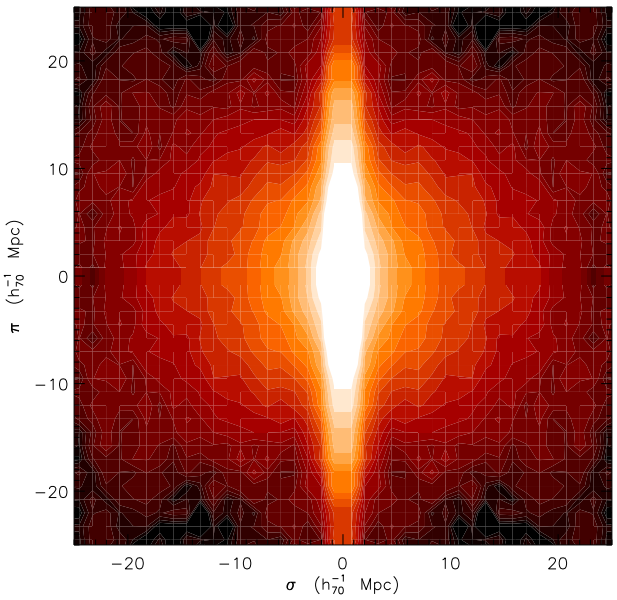

Figure 4.16 The two-dimensional correlation function, $\xi(\sigma, \pi)$, of the SDSS "red" sample (see Fig. 4.3). The separation bins and contour levels are the same as for Fig. 4.14.

effect (i.e. the elongated structure along the π-axis at $\sigma \approx 0$ Mpc). By contrast, the red SDSS subsample shows a $\xi(\sigma, \pi)$ with much larger overall amplitude, as well as a very distinct finger of god feature. In addition, the $\xi(\sigma, \pi)$ contours for the red SDSS sample display a more symmetric, "round", shape on intermediate scales ($\gtrsim 10$ Mpc).

These results are not unexpected; it is well established that gas-rich galaxies are associated with late-type morphology, blue optical colors and elevated specific star formation rates [e.g. Huang et al., 2012a, Catinella et al., 2010, Li et al., 2012a]. For example, Huang et al. [2012a] shows that the ALFALFA sample is heavily biased against red-sequence galaxies, while sampling very well the less luminous and more actively star-forming galaxies galaxies in the "blue cloud" (their Fig. 11). The main conclusions summarized in points 1-3, therefore, are a direct consequence of the fact that blue galaxies are significantly less clustered than red galaxies, irrespective of luminosity [see e.g. Fig 16 in Zehavi et al.,

2011]. The bias of blind HI surveys against red-sequence galaxies also helps explain the marked difference in the shape of $\xi(\sigma, \pi)$ between the ALFALFA and SDSS red samples (Figs. 4.14 & 4.16, respectively). Red galaxies are usually found in high density environments, such as clusters of galaxies and compact groups, and their clustering bears the signs of the large peculiar velocities found in these environments. In particular, the red sample has an increased number of galaxy pairs that have small physical but large velocity separations; these pairs produce the strong "finger of god" feature in $\xi(\sigma, \pi)$ at $\sigma \approx 0$. On the other hand, galaxies with blue colors and HI galaxies tend to inhabit the lower density "field". As a result, they trace the ordered flow towards matter overdensities without significant noise from peculiar motions. This is why the characteristic asymmetric shape of $\xi(\sigma, \pi)$ at separations $\gtrsim$10 Mpc, which is caused by these systematic motions, is more pronounced in the blue and HI samples.

4.3.3 Cross-correlation between HI-selected and optically selected samples

The results above can be used to compare the clustering properties of HI and optical galaxies, but do not contain information about the spatial relationship among the samples under consideration. In particular, they cannot address questions such as whether or not HI galaxies inhabit the same environments as a given class of optical galaxies. It is already known through the study of individual clusters [Giovanelli and Haynes, 1985, Haynes and Giovanelli, 1986, Solanes et al., 2002] that galaxies in high density environments tend to have lower gas fractions than their counterparts in the field, and thus have a lower

probability of being included in an HI-selected sample. Since optical galaxies with red colors are found preferentially in dense environments, we expect HI-selected samples to show some degree of "segregation" with respect to red galaxies. Here, we use the large galaxy samples provided by the ALFALFA and SDSS surveys to obtain a statistical measurement of this effect, and to pin down the length scale over which environment can affect the gaseous contents of galaxies.

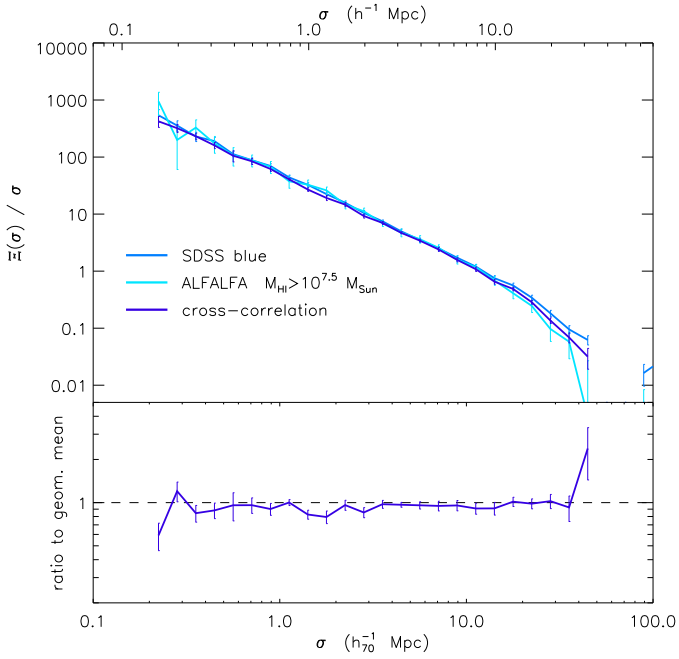

Figure 4.17 *upper panel:* The projected correlation function, $\Xi(\sigma) / \sigma$, for the blue SDSS sample (dark blue solid line) and the parent ALFALFA sample (light blue solid line), compared to their projected cross-correlation function (purple solid line). *lower panel:* The ratio of the cross-correlation function to the geometric mean of the correlation functions of the two constituent samples, $\mathcal{R}(\sigma)$.

The spatial relationship of two galactic samples is encoded in their cross-correlation function. In this chapter, we calculate cross-correlation functions using a modified version of the LS estimator [following Zehavi et al., 2011]:

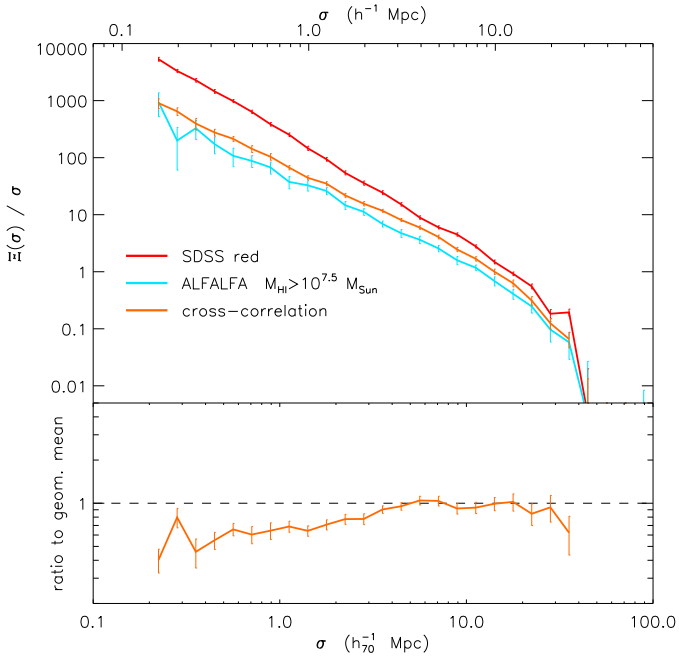

Figure 4.18 *upper panel:* The projected correlation function, $\Xi(\sigma) \,/\, \sigma$, for the red SDSS sample (red solid line) and the parent ALFALFA sample (light blue solid line), compared to their projected cross-correlation function (orange solid line). *lower panel:* The ratio of the cross-correlation function to the geometric mean of the correlation functions of the two constituent samples, $\mathcal{R}(\sigma)$. Note the clear tendency for SDSS galaxies with red colors and HI-rich ALFALFA galaxies to avoid each other at separations $\lesssim 3$ Mpc.

$$\hat{\xi}_{cross} = (DD_1DD_2 - DD_1RR_2 - DD_2RR_1 + RR_1RR_2)/RR_1RR_2 \ . \qquad (4.15)$$

Here DD, RR and DR are the normalized data-data, random-random and data-random pair counts, and the subscripts 1 & 2 are used to denote the two samples. Generally, the information present in the cross-correlation function is most intuitively presented in terms of its ratio with the geometric mean of the correlation functions of the two constituent samples, $\mathcal{R}(r) \equiv \xi_{cross}(r)/\sqrt{\xi_1(r)\xi_2(r)}$. In essence, $\mathcal{R}$ measures the degree to which two samples are spatially "aware" of one another: Two spatially independent samples have $\mathcal{R} = 1$ for all r (i.e. $\xi_{cross}(r) = \sqrt{\xi_1(r)\xi_2(r)}$). Conversely, a ratio of $\mathcal{R} < 1$ at some separation r means that the two samples "avoid" each other on the length scale under consideration.

Figure 4.17 shows the cross-correlation function between the ALFALFA parent sample ($M_{HI} > 10^{7.5}\ M_{\odot}$) and the blue SDSS sample. The lower panel shows that $\mathcal{R} \approx 1$ on all probed scales, meaning that HI-selected galaxies and optical galaxies with blue colors have no special spatial relationship. In other words, detecting a blue SDSS galaxy at a given location in space does not influence our chance of finding an ALFALFA-detected galaxy in its vicinity, beyond what is expected from the clustering of the samples. The situation is very different in Figure 4.18, which shows that the cross-correlation function between HI-selected galaxies and optical galaxies with red colors is systematically lower than their geometric mean at small separations (i.e. $\mathcal{R} < 1$ at $\sigma \lesssim 3$ Mpc). This means that the existence of a red SDSS galaxy at a given position in space lowers the chances that an HI-rich galaxy is positioned within ≈ 3 Mpc from it.

These results also hold for the two-dimensional cross-correlation functions

between the HI-selected ALFALFA sample and the color-based SDSS samples. Figure 4.19, for example, shows a two-dimensional map of $\mathcal{R}(\sigma, \pi)$ calculated from the cross-correlation function between the ALFALFA parent sample and the SDSS blue sample. Regions that are enclosed by solid contours are those where $\mathcal{R}$ deviates significantly from unity ($\mathcal{R} < 0.85$ or $\mathcal{R} > 1.15$). We can clearly see that, barring the large fluctuations at the outskirts of the map caused by noise, $\mathcal{R}(\sigma, \pi) \approx 1$ over most of the map. The situation is very different when the cross-correlation between the ALFALFA parent sample and the SDSS red sample is considered. Figure 4.20 shows that regions corresponding to $\sigma \lesssim 3$ Mpc have systematically low values of $\mathcal{R}$, over the whole range of π-axis separations. This characteristic shape demonstrates graphically that HI-selected galaxies avoid regions of space where the finger-of-god effect is large, corresponding mostly to galaxy clusters and rich groups.

This measurement of the cross-correlation properties of HI galaxies with respect to various optical samples is especially important in the context of cosmological studies with next generation HI surveys [e.g. Beutler et al., 2011], and 21cm intensity-mapping experiments at moderate redshift [e.g. Masui et al., 2013]. In particular, Figs. 4.11 - 4.20 show that an HI-selected sample traces the cosmic large-scale structure differently than most optical surveys. For example, due to the very different clustering properties of HI-rich galaxies and galaxies with red colors, a 21cm survey would provide a very different view of the large scale structure compared to a survey of, e.g., luminous red galaxies [as in e.g. Eisenstein et al., 2005]. On the other hand, a survey targeting actively star-forming galaxies (such as the UV-selected WiggleZ survey; Drinkwater et al., 2010) will be a much closer match in terms of clustering properties and spatial distribution. The considerations above have an effect on the potential of future

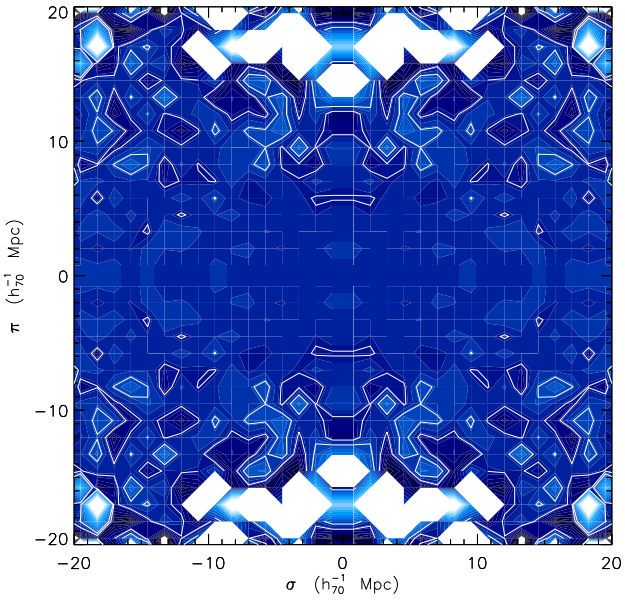

Figure 4.19 Two-dimensional map of $\mathcal{R}(\sigma, \pi)$ for the ALFALFA parent sample ($M_{HI} > 10^{7.5}\ M_{\odot}$) and the SDSS blue sample. $\mathcal{R}(\sigma, \pi)$ is the ratio of the cross-correlation function between the two samples to the geometric mean of their respective correlation functions. The separation bins are the same as in Fig. 4.14. The contour levels are logarithmically spaced, with values doubling every six contours. The darkest shade corresponds to the minimum value of 0.25 while the lightest shade corresponds to the maximum value of 4. Regions enclosed by solid contours are regions where $\mathcal{R}$ deviates significantly from unity ($\mathcal{R} < 0.85$ or $\mathcal{R} > 1.15$).

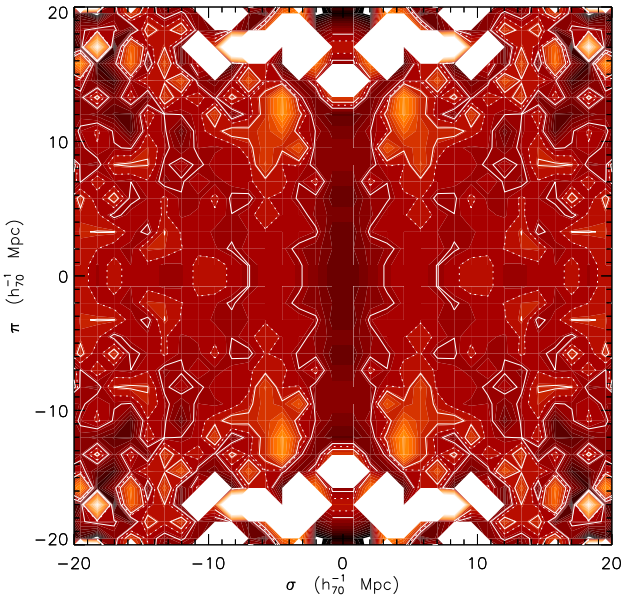

Figure 4.20 Same as Fig. 4.19 but regarding the cross-correlation of the ALFALFA parent sample ($M_{HI} > 10^{7.5}\ M_{\odot}$) with the SDSS red sample. The separation bins and contour levels are also the same as in Fig. 4.19. Regions enclosed by solid contours are regions where $\mathcal{R}$ deviates significantly from unity ($\mathcal{R} < 0.85$ or $\mathcal{R} > 1.15$). Since the central part of the map contains systematically low values of $\mathcal{R}$, an additional dotted contour level at $\mathcal{R} = 1$ has been drawn to lift ambiguities.

21cm surveys for cosmological studies: For example, the low clustering amplitude of HI galaxies may lower our sensitivity for detecting the BAO feature in the large-scale galaxy correlation function. On the other hand, measurements that are based on the anisotropy of $\xi(\sigma, \pi)$ may benefit considerably from the low levels of peculiar motion "noise" achieved with an HI-selected sample. As a result, HI surveys may prove advantageous in measuring redshift-space distortions (RSD) and the growth of structure ("$f\sigma_8$" measurements, e.g. Reid et al., 2012, de la Torre et al., 2013, Contreras et al., 2013).

4.4 Which halos host gas-rich galaxies?

4.4.1 ΛCDM halo sample

We select a sample of dark matter halos from the Bolshoi ΛCDM simulation [Klypin et al., 2011]. The Bolshoi simulation is a high-resolution dissipationless simulation, run for a set of cosmological parameters consistent with the 7-year results of the *Wilkinson Microwave Anisotropy Probe* [WMAP; Jarosik et al., 2011] and other recent cosmological studies. We use the halo catalogs[7] extracted with the Bound-Density-Maxima (BDM) halo finder algorithm [Klypin and Holtzman, 1997]. In particular, we select a box region of the Bolshoi simulation of size $\approx 140\ h_{70}^{-1}$Mpc on a side, such that the volume of the halo sample is comparable to the ALFALFA volume. In addition, we restrict ourselves to halos with maximum circular velocities $v_{halo} > 60\ \mathrm{km\,s^{-1}}$, corresponding to $M_{vir} \gtrsim 10^{11}\ M_{\odot}$. In total, our halo sample consists of $94\,671$ halos, including both distinct halos as well as subhalos. From this parent halo sample we create subsamples by spec-

ifying five v_{halo} ranges, as shown in Figure 4.21: 60-82 $\mathrm{km\,s^{-1}}$, 82-114 $\mathrm{km\,s^{-1}}$, 114-157 $\mathrm{km\,s^{-1}}$, 157-217 $\mathrm{km\,s^{-1}}$ and >217 $\mathrm{km\,s^{-1}}$.

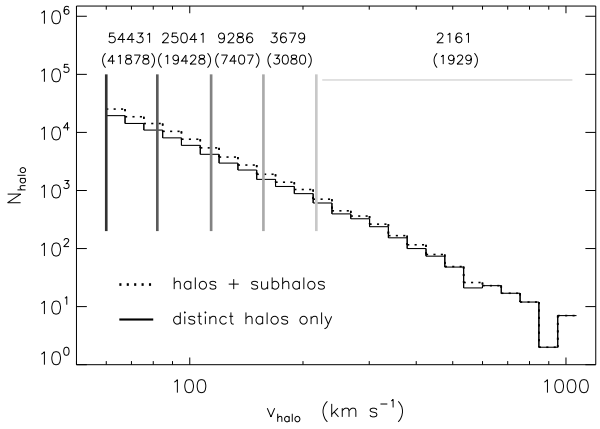

Figure 4.21 Histogram of maximum circular velocities, v_{halo}, for halos selected from the Bolshoi ΛCDM simulation [Klypin et al., 2011]. The solid histogram represents the counts of distinct halos only, while the dotted histogram refers to all halos (including both distinct halos as well as subhalos). The vertical solid lines denote the boundaries of the five velocity-binned samples, described in §4.4.1. The quoted numbers correspond to the overall halo count in each sample (including halos & subhalos), while the numbers in parentheses denote the number of distinct halos only. Note that, unlike the similar Figs. 4.1 & 4.2, both axes in this Figure are logarithmic.

Furthermore, we consider halo subsamples split by their spin parameter, λ_K, defined as:

$$\lambda_K = \frac{J_{vir}\, K^{1/2}}{G\, M_{vir}^{5/2}} \quad . \tag{4.16}$$

Note that the BDM database for the Bolshoi simulation reports spin parameters defined in terms of the halo kinetic energy (K), instead of the more common definition based on total energy ($\lambda = J_{vir}|E_{tot}|^{1/2}/GM_{vir}^{5/2}$). However, the two definitions yield very similar results for well-virialized halos, since in this case $K \approx |E_{tot}|$. Figure 4.22 displays graphically the three spin-based subsamples,

[7] `www.multidark.org/MultiDark/Help?page=databases/bolshoi/database`

referred to as “low spin”, “average spin” and “high spin” samples.

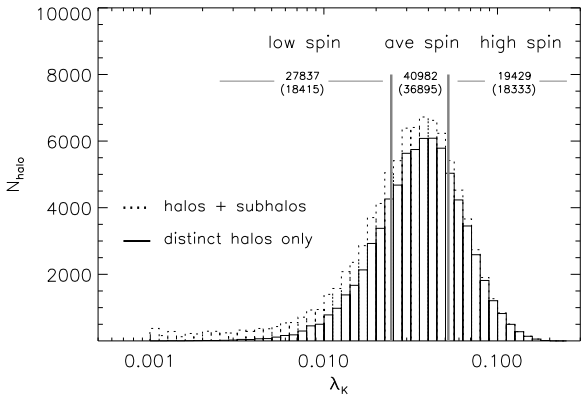

Figure 4.22 Histogram of “kinetic” spin parameter, λ_K (see Eqn.4.16 for definition), for halos selected from the Bolshoi ΛCDM simulation [Klypin et al., 2011]. The solid histogram represents the counts of distinct halos only, while the dotted histogram refers to all halos (including both distinct halos as well as subhalos). The vertical solid lines denote the boundaries of the three spin-based samples. The quoted numbers correspond to the overall halo count in each sample (halos & subhalos), while the numbers in parentheses denote the number of distinct halos only.

The halo samples described above are volume-limited, meaning that inclusion in some specific sample does not depend on their position in the simulation box. As a result, all halo samples share the same random catalog, which is straightforwardly created by a set of points with uniformly distributed x, y, z coordinates. In addition, no pair-weighting (see §4.2.5) is necessary, since all halo pairs in the simulation can be accounted for. Lastly, the separations between pairs of halos are readily available in terms of physical length, and are not affected by redshift-space distortions. It follows that for these halo samples $\xi(r) \equiv \xi(s)$, while $\Xi(\sigma) \,/\, \sigma$ can be calculated by projecting separations on any arbitrary axis.

4.4.2 Halo mass & halo/subhalo status

Figures 4.23 & 4.24 show the projected correlation functions of the velocity-binned halo samples, including and excluding subhalos from the computation respectively. Overplotted on both Figures is the projected correlation function of the $M_{HI} > 10^9\ M_\odot$ ALFALFA sample (solid cyan line), for comparison purposes. In both Figures there is a clear trend for more massive halos to show increasingly stronger clustering. This trend is consistent with theoretical expectations [e.g. Musso et al., 2012], since more massive halos are expected to form in regions with larger matter overdensity. This behavior is not shared by the HI mass-thresholded samples in this work, which do not display any discernible clustering enhancement with increasing HI mass (see Figure 4.8). This fact alone suggests that galaxy HI mass is not tightly related to the mass of the host halo. Furthermore, a comparison of Figures 4.23 & 4.24 shows that the inclusion of subhalos in a sample leads to higher amplitude clustering, especially at small separations. This is also expected, since subhalos are found in the vicinity of other halos by definition. More specifically Figure 4.23 shows that, when subhalos are included, all halo samples display stronger clustering than ALFALFA galaxies. This further suggests that a sizable fraction of subhalos do not host HI galaxies. In particular, according to Figure 4.24, if HI galaxies were exclusively hosted by distinct halos then the clustering of typical ALFALFA galaxies would be best matched by the clustering of halos with $v_{halo} \approx$ 100-150 $\mathrm{km\,s^{-1}}$. In the more realistic intermediate case, where some but not all subhalos host HI galaxies, typical ALFALFA galaxies would be hosted by slightly less massive halos ($v_{halo} \lesssim 100\ \mathrm{km\,s^{-1}}$).

Alternatively, we can use the technique of abundance matching to study the

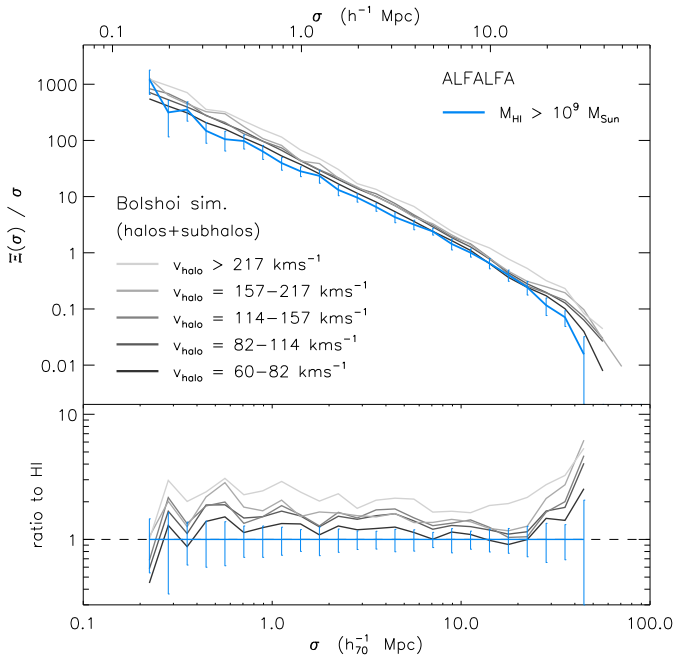

Figure 4.23 *upper panel:* The projected correlation function for the velocity-binned Bolshoi halo samples (shown in Figure 4.21 and described in §4.4.1), compared to the correlation function of the $M_{HI} > 10^9\ M_\odot$ ALFALFA sample. Both halos & subhalos are included in the computation of the halo $\Xi(\sigma)\,/\,\sigma$. Darker shades represent samples a lower v_{halo} range, as depicted in the Figure legend. *lower panel:* The same projected correlation functions as above, normalized to the correlation function of the ALFALFA sample. The unity line is also plotted for reference.

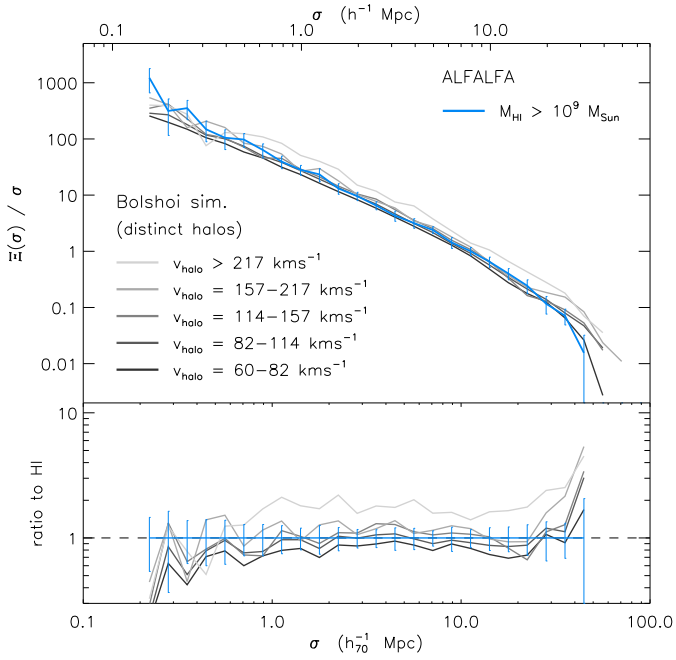

Figure 4.24 Same as in Fig. 4.23, but for distinct halos only.

relation between our HI-selected sample and their host dark matter halos in a more systematic way. Abundance matching is a simple, yet powerful statistical approach to connect galaxy properties (such as the luminosity, stellar or baryonic mass, velocity, etc.) to their host dark matter (sub)halos [e.g., Shankar et al., 2006, Conroy et al., 2006, Guo et al., 2010, Reddick et al., 2012, Rodríguez-Puebla et al., 2012, Papastergis et al., 2012]. In its most simple form, the observed abundances of galaxies at a given property are matched against the theoretical halo plus subhalo abundances. The result is a galaxy property versus halo mass empirical relation. In reality, galaxy properties are not determined solely by the mass of the halo in which they reside but, due the complexity of the galaxy formation process, a dependence on other halo and/or environmental properties is expected. To take this into account, recent works have extended the abundance matching technique to include a scatter around the mean relation [e.g., Behroozi et al., 2010, Moster et al., 2010, Hearin et al., 2012, Rodríguez-Puebla et al., 2013].

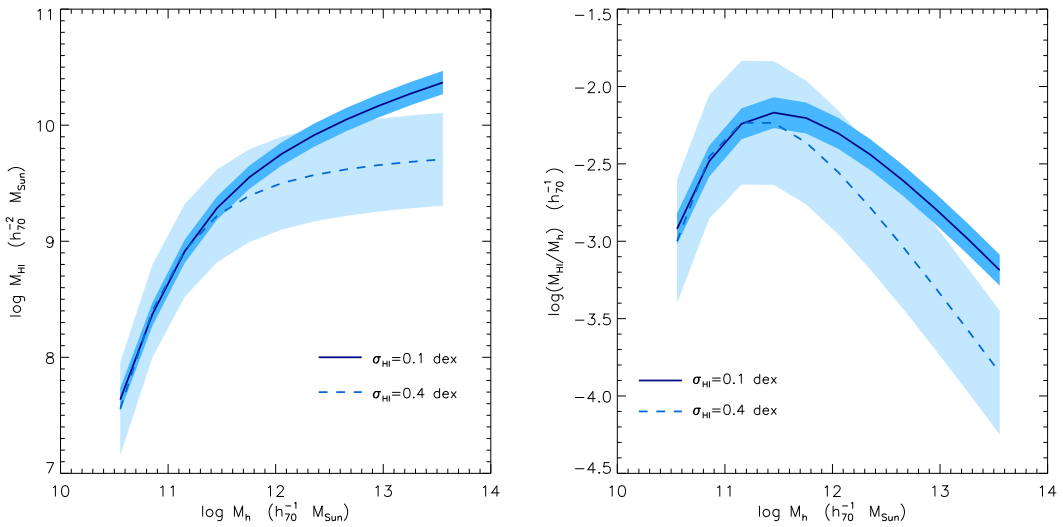

Figure 4.25 *Left panel:* The solid and dashed lines represent the M_{HI}-M_h relation obtained via abundance matching, assuming a scatter of $\sigma_{HI} = 0.1$ dex and $\sigma_{HI} = 0.4$ dex, respectively. The shaded areas denote the assumed scatter around each average relation. Note that, in the $\sigma_{HI} = 0.4$ dex case, the HI mass is nearly independent of halo mass for $M_{HI} \gtrsim 10^{9.5}\ M_\odot$. *Right panel:* The solid and dashed curves represent the M_{HI}/M_h ratio as a function of M_h in the $\sigma_{HI} = 0.1$ dex and $\sigma_{HI} = 0.4$ dex cases, respectively. Note that in both cases, the maximum M_{HI}/M_h value is reached at relatively low halo masses ($M_h \approx 10^{11.3}\ M_\odot$ and $M_h \approx 10^{11.1}\ M_\odot$, respectively).

Here we use the abundance matching technique in order to obtain an average relation between galaxy HI mass and host (sub)halo mass (M_{HI}-M_h relation). To do so, we employ the observed HI mass function of the HI-selected sample of Papastergis et al. [2012, see their Table 1] and the halo plus subhalo mass function obtained from the Bolshoi simulation (see §4.4.1). Reddick et al. [2012] have shown that the measure of halo mass that is most tightly related to the stellar properties of galaxies (e.g. stellar mass or optical luminosity) is the maximum mass reached along the entire merger history of the (sub)halo, $M_{h,\rm peak}$; we therefore perform our abundance matching analysis using $M_{h,\rm peak}$ as the halo mass, dropping from now on the excess notation ($M_{h,\rm peak} \rightarrow M_h$). Note that $M_{h,\rm peak}$ is approximately equal to the present-day mass for distinct halos, but in the case of subhalos it can be significantly larger than their present-day mass, due to the effects of tidal stripping.

We furthermore assume that the distribution of HI mass at a given (sub)halo mass is drawn from a lognormal distribution with mean $M_{HI}(M_h)$ and a scatter of σ_{HI} around it. Here we will assume that σ_{HI} is independent of halo mass. While the scatter around the average stellar mass-halo mass relation ($M_*(M_h)$ relation) has been discussed extensively in the literature [e.g., Cacciato et al., 2009, More et al., 2009, Yang et al., 2009, More et al., 2011, Leauthaud et al., 2012, Rodríguez-Puebla et al., 2013], σ_{HI} has not been discussed previously. Here we opt to use two different values for σ_{HI} in order to gauge the uncertainty introduced from our lack of knowledge on its value and mass dependence: $\sigma_{HI} = 0.1$ dex and $\sigma_{HI} = 0.4$ dex. A more thorough exploration of this scatter is deferred for a future publication.

The left-hand panel of Figure 4.25 shows the resulting average $M_{HI}(M_h)$ re-

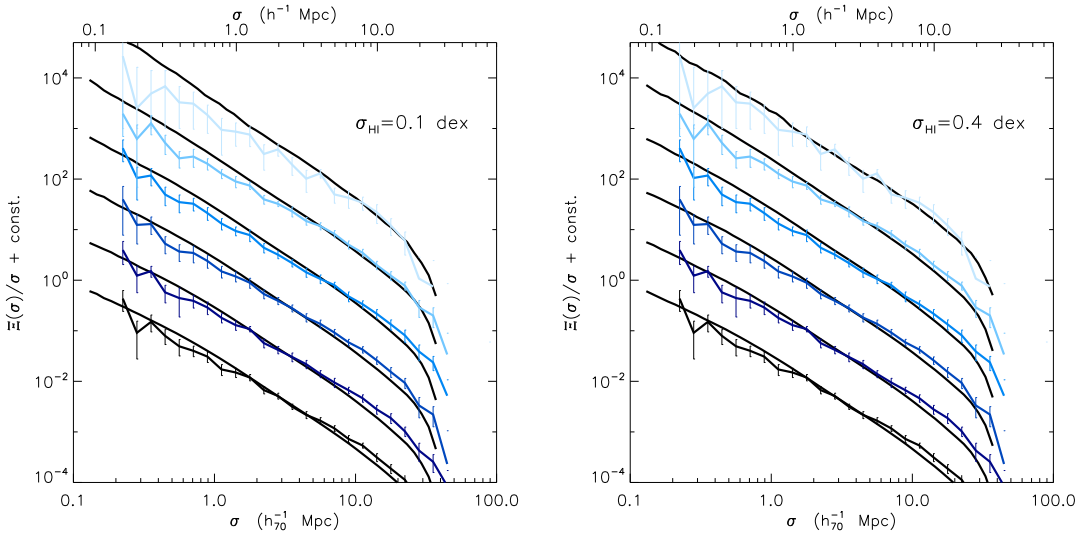

Figure 4.26 Comparison of the correlation functions of the ALFALFA HI mass-thresholded samples with simulated samples obtained by assigning M_{HI} values to each Bolshoi (sub)halo according to the relation in Fig. 4.25. Each ALFALFA sample is plotted in a different shade of blue, while all simulated samples are plotted as black solid lines. From top to bottom each pair of lines corresponds to the $M_{HI} > 10^{10}, 10^{9.5}, 10^{9}, 10^{8.5}, 10^{8}, 10^{7.5}$ $M_{\odot}$ samples (a constant offset between samples has been used for clarity). Note that the lower ends of some error bars for the $M_{HI} > 10^{10}$ $M_{\odot}$ ALFALFA sample have been clipped for clarity. *Left panel:* M_{HI} values are assigned to each Bolshoi (sub)halo via abundance matching assuming $\sigma_{HI} = 0.1$ dex (solid line in Fig. 4.25). *Right panel:* Same as left panel, but assuming $\sigma_{HI} = 0.4$ dex (dashed line in Fig. 4.25). Observations seem to favor a larger value of scatter.

lations for both values of σ_{HI}. In both cases the HI mass dependence with halo mass raises steeply at low and intermediate halo masses ($M_h < 10^{11.5} M_\odot$), scaling roughly as $M_{HI} \propto M_h^{1.8}$. For high halo masses the average $M_{HI}(M_h)$ relation becomes shallower, while in the $\sigma_{HI} = 0.4$ dex case M_{HI} it is nearly independent of halo mass. According to the abundance matching result, galaxies with $M_{HI} > 10^9\ M_\odot$ are hosted by halos with $M_h \gtrsim 10^{11.3}\ M_\odot$, or $v_{halo} \gtrsim 90\ \mathrm{km\,s^{-1}}$, in reasonable agreement with our claims based on Figures 4.23 & 4.24. The right-hand panel of Figure 4.25 shows the average M_{HI}/M_h ratio as a function of M_h. The maximum of the HI-to-halo mass ratio is obtained at $M_h \approx 10^{11.3} M_\odot$ in the case of $\sigma_{HI} = 0.1$ dex, and at $M_h \approx 10^{11.1} M_\odot$ in the $\sigma_{HI} = 0.4$ dex case. Note that both values are lower than the values commonly obtained for the location of the peak of the M_*/M_h ratio, which is $M_h \approx 10^{12} M_\odot$.

Once we have assigned a value of M_{HI} to each (sub)halo of the Bolshoi simulation, we can compute the correlation functions of modeled samples with any range of HI mass, and compare them with the ALFALFA results. For example, in Figure 4.26 we compare the projected correlation functions of the ALFALFA HI mass-thresholded samples to the correlation functions of modeled samples with the same HI thresholds. Overall, we find that the clustering dependence on HI mass is rather weak. Nevertheless, in the $\sigma_{HI} = 0.1$ dex case galaxies with large HI masses ($M_{HI} > 10^{10} M_\odot$) are expected to show stronger clustering than lower HI mass samples. On the other hand, in the $\sigma_{HI} = 0.4$ dex case the clustering amplitude of galaxies is expected to be almost independent of HI mass. This latter case is therefore in better agreement with the observational results of Fig. 4.8. We conclude that the clustering properties of ALFALFA galaxies favor a large scatter in the M_{HI}-M_h relation, and a weak dependence of galactic HI mass on host halo mass (at least for galaxies with $M_{HI} \gtrsim 10^{9.5}\ M_\odot$).

Moreover, Figure 4.26 shows that all modeled samples display consistently stronger clustering than the actual ALFALFA samples. This is because our abundance matching analysis assumes that all subhalos host an HI galaxy. If we repeat the analysis by considering only distinct halos we find the opposite result, i.e. that all modeled samples consistently underestimate the clustering amplitude of the actual ALFALFA samples (Figure not shown). Our second conclusion is therefore that only a subset of subhalos host HI galaxies, with the rest presumably hosting gas-poor galaxies that are not detected by ALFALFA. In view of the results above, it is important to ask whether halo properties other than mass and halo/subhalo status may be playing a major role in determining the gas content of galaxies.

4.4.3 Halo spin parameter

The spin parameter of the host halo has been suggested to be the decisive factor in setting a number of galaxy properties, such as the the galaxy's stellar and gas surface density. In fact, low surface brightness (LSB) galaxies are currently believed to be hosted by halos with higher-than-average spin parameters [e.g. Boissier and Prantzos, 2000]. Several lines of evidence also suggest that halo spin may be closely related to the overall gas-to-stellar mass ratio of a galaxy, in the sense that halos with higher spin parameters host more gas-rich systems at fixed stellar mass. Firstly, gas-rich galaxies are known to be of relatively low surface brightness and low stellar mass density [Catinella et al., 2010, Zhang et al., 2009, Huang et al., 2012a], properties that are typically associated with high spin halos. Huang et al. [2012a] have furthermore directly estimated the galactic spin parameter of the entire ensemble of ALFALFA galaxies, obtaining

the result that measured galaxy spin increases with increasing gas-fraction (their Figure 14). Lastly, recent hydrodynamical simulations by Kim and Lee [2012] have shown that, at fixed halo mass, halos with higher spin parameters have more extended gaseous disks and larger overall gas-to-stellar mass ratios (Jihoon Kim, private communication).

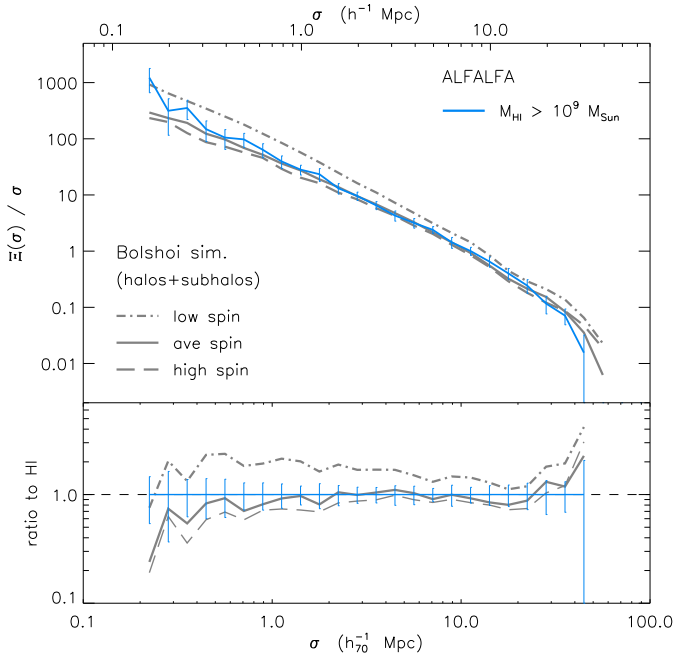

Figure 4.27 *upper panel:* The projected correlation function for the Bolshoi halo samples split by spin (shown in Figure 4.22 and described in §4.4.1), compared to the correlation function of the $M_{HI} > 10^9\ M_\odot$ ALFALFA sample (cyan line). Both halos & subhalos are included in the computation of the halo $\Xi(\sigma)\,/\,\sigma$. The solid, dot-dashed and long-dashed grey lines represent the correlation functions of the "average spin", "low spin" and "high spin" samples, respectively. *lower panel:* The same projected correlation functions as above, normalized to the correlation function of the ALFALFA sample. The unity line is also plotted for reference.

Here we aim to investigate the spin-gas content relation by comparing the clustering properties of the gas-rich ALFALFA galaxies and of halos with "low", "average" and "high" spin parameters. Figure 4.27 shows the projected correlation functions for the three halo samples, and compares them with the projected correlation function of the $M_{HI} > 10^9\ M_\odot$ ALFALFA sample (solid cyan

line). Note that both distinct halos and subhalos are included in the computation of $\Xi(\sigma) \,/\, \sigma$ for these halo samples, and no cuts are performed based on halo velocity. The Figure clearly shows that halos with low spin parameters display a markedly stronger correlation function compared to halos with average and high spin parameters. This behavior can be attributed to two main causes: Firstly, the "low" spin sample contains proportionally more subhalos than the "average" and "high" spin samples (see Fig.4.22), which tend to be a highly clustered population. Secondly, low spin parameter halos tend to cluster significantly more than their higher spin counterparts, even when the sample is restricted to distinct halos only (plot not shown). Note that the results are robust[8] under different definitions of the halo spin parameter, and hold also for the "classical" spin parameter, defined as $\lambda = J\,|E_{\rm tot}|^{1/2} \,/\, G\,M_h^{5/2}$. Regardless of the underlying reason, the fact that a crude cut on halo spin parameter alone can reproduce fairly well the correlation function of ALFALFA galaxies is remarkable, and lends further support to the hypothesis that halo spin parameter –and not halo mass– is the main property setting the gas content of galaxies.

This result could also have important implications for the modeling of gas-rich galaxies, which will be an integral part of the scientific interpretation of near-future 21cm large-scale surveys. In particular, it may be necessary for semi-

[8]We have also carried out this analysis with one additional halo catalog for the Bolshoi simulation, extracted by the ROCKSTAR halo finder [Behroozi et al., 2013a] and refined according to Behroozi et al. [2013b] (`www.slac.stanford.edu/~behroozi/Bolshoi_Catalogs/`). In contrast with BDM halos, ROCKSTAR halos do *not* display clustering variations when split based on their spin parameter. At the same time, ROCKSTAR halos have a different distribution of spin parameters than BDM halos, with the most conspicuous difference being the lack of the low-spin tail in the distribution of ROCKSTAR subhalos (see Fig. 4.22). The large difference in the spin measurements of the two halo finders seems to be caused by the fact that BDM measurements are made in spherical shells, while ROCKSTAR measurements can be performed over significantly asymmetric domains, both in real and velocity space (Anatoly Klypin, private communication). In this chapter we choose to focus on results obtained with the BDM halo finder, because spin measurements with BDM are weighted more towards the inner regions of halos, which may better correspond to the spin of the baryonic component hosted by the halos.

analytic models (SAMs) of HI galaxies [e.g. Obreschkow et al., 2009, Lu et al., 2012] to consider halo spin –in addition to mass– in their implementation. Empirical approaches such as abundance matching may also need to be revised, and may potentially need to consider matching HI galaxy abundances with the joint distribution of halo mass and spin [see e.g. Hearin and Watson, 2013 for a similar approach pertaining to galaxy optical colors].

4.5 Conclusion

We use the sample of galaxies detected by the ALFALFA blind 21cm survey, to study the clustering characteristics of HI-selected galaxies (i.e. galaxies selected based on their atomic hydrogen content). In particular, we divide the ALFALFA galaxies into subsamples based on their HI mass, creating six HI mass-thresholded and three HI mass-binned samples, spanning the entire $M_{HI} = 10^{7.5} - 10^{11}\ M_{\odot}$ range. We measure the projected correlation function for each of the samples above, and find no compelling evidence for a dependence of clustering on HI mass. The data does yield a lower amplitude correlation function for the least massive HI mass-binned sample ($M_{HI} = 10^{8.5} - 10^{9.5}\ M_{\odot}$), but we attribute this effect to the small volume probed by the specific sample (see Figures 4.9 & 4.10).

Moreover, we compare the clustering characteristics of the HI sample with those of optically selected galaxies drawn from the SDSS spectroscopic database. We follow a similar procedure as described above, and divide the optical galaxies into subsamples based on thresholds and ranges on their r-band absolute magnitude (see Figure 4.2). In addition, we split the parent optical

sample into three color-based subsamples, based on the galaxies' position on a color-magnitude diagram (see Figure 4.3). We find that HI-selected galaxies cluster more weakly than their optical counterparts, even those at faint absolute magnitudes (at least as faint as $M_r \approx -18$). On the other hand, we find that the correlation function of ALFALFA galaxies is matched extremely well by the correlation function of SDSS galaxies with blue colors. Conversely, SDSS galaxies with red colors display much stronger clustering than the HI-selected samples, resulting in a projected correlation function with a markedly steeper slope and higher amplitude. The results above hold also for the full two-dimensional correlation function, $\xi(\sigma, \pi)$: both the ALFALFA and SDSS blue samples display a strongly anisotropic shape at scales $\gtrsim$10 Mpc, and a very weak "finger of god" feature at small on-sky separations (see Figures 4.14 & 4.15). On the other hand, the two-dimensional correlation function of SDSS red galaxies shows a prominent finger of god feature and a more isotropic shape at intermediate scales (Figure 4.16).

In addition, we carry out a cross-correlation analysis between the ALFALFA and color-based SDSS samples. The HI$\times$red cross-correlation function shows that the gas-rich ALFALFA galaxies "avoid" being located within $\approx$3 Mpc of optical galaxies with red colors. In particular, they avoid environments where the finger of god effect is strong, presumably corresponding to clusters and rich groups. This amounts to a statistical measurement of the "HI deficiency" of galaxies in clusters, and yields a quantitative result for the length scale over which dense environments typically affect the gas contents of galaxies.

We also measure the clustering properties of the halos in the Bolshoi ΛCDM simulation, to gain insights on the characteristics of halos hosting gas-rich galax-

ies. By comparing the clustering of ALFALFA galaxies to that of halo samples split based on their mass and halo/subhalo status we arrive at the conclusions that i) HI mass is not tightly related to the mass of the host halo and ii) a sizable fraction of subhalos does not host gas-rich galaxies. We furthermore perform a more detailed modeling of the clustering of halos hosting gas-rich galaxies, based on the M_{HI}-M_h relation inferred from abundance matching. The results confirm our previous findings, by favoring an M_{HI}-M_h relation that has large scatter and a weak dependence of M_{HI} on host halo mass (see Figures 4.25 & 4.26). Lastly, we consider the consider the clustering of halos with different spin parameters. We find that halos with low spin parameters (as measured by the Bound-Density-Maxima halo finder algortihm) cluster more strongly than halos with higher spin parameters. Remarkably, this leads to the correlation function of ALFALFA galaxies being reproduced fairly well by just excluding low-spin halos from the computation (Figure 4.27). This finding provides indirect support to the hypothesis that halo spin plays a central role in determining the gas contents of present-day galaxies.

CHAPTER 5

OVERVIEW & CONCLUSIONS

The work presented in this dissertation focuses on the statistical analysis of the galactic sample detected by the Arecibo Legacy Fast ALFA (ALFALFA) survey, and the interpretation of the results with regards to the properties of dark and baryonic matter on the scale of galaxies.

Until recently, such analyses were reserved for the samples of large optical surveys, such as the Two-degree Field Galaxy Redshift Survey (2dFGRS) and the Sloan Digital Sky Survey (SDSS), who have recorded $>$1 million optical spectra and $\sim$tens of millions of photometric optical sources. Along with an explosion in data, came an explosion of literature studies devoted to the statistical analysis of galaxy ensembles. Topics range from measurements of the abundance of galaxies as a function of their luminosity (i.e. galaxy luminosity function), to their color distribution (i.e. color-magnitude diagram), to their clustering properties on both small and large scales. These works have been of fundamental importance for our understanding of how galaxies form and evolve, advancing tremendously the knowledge gained over the previous decades from object-by-object studies. In fact, the volumes probed by optical surveys today are large enough, that measurements of galaxy statistical properties are now a standard tool of observational cosmology as well.

The ALFALFA survey marks the beginning of an era where blind HI surveys produce samples that are large enough to undertake detailed statistical studies of the galaxies in the local universe. The presently available public ALFALFA catalog ("α.40" catalog) contains the first sample with $> 10\,000$ blindly-detected 21cm galaxies. When source extraction will be completed, ALFALFA will have

detected $> 30\,000$ sources, almost an order of magnitude more than its very successful predecessor, the HI Parkes All-Sky Survey (HIPASS). In addition to the largest HI-selected galaxy sample to date, ALFALFA can claim a wide sky coverage ($\approx 6\,300$ deg^2 for the completed survey) and superior sensitivity and spectral resolution (can probe the galaxy population robustly down to $M_{HI} \sim 10^7\ M_\odot$ and $v_{rot} \sim 20\ \mathrm{km\,s^{-1}}$).

All of these characteristics allow one to undertake statistical studies of 21cm galaxies at a level of detail not previously achieved. These studies, in turn, inform our theories of galaxy formation and evolution by offering data on a critical –but largely overlooked– galactic component: their gas content. Gas content is important not only because it determines a galaxy's potential for future star formation, but also because it is the most immediate "victim" of stellar feedback. Given that the implementation of stellar feedback is one of the most important and least constrained aspects of hydrodynamic simulations of galaxy formation, measurements of the gas contents of galaxies have the potential to test stringently our models of galaxy formation.

Due to their intrinsic spectroscopic nature, 21cm surveys have also the ability to carry out investigations that are not amenable to optical techniques. For example, the ALFALFA survey obtains automatically the global 21cm line profile of every detected source. One can therefore use the ALFALFA dataset to obtain information about the internal kinematics of thousands of nearby galaxies, as encoded in each one's spectral profile width and shape.

Below follows a brief summary of the work presented in this dissertation, including the main results of the studies presented in Chapters 2, 3 & 4. This dissertation reflects the range of research questions that can be addressed by

wide-area blind HI surveys. With the completion of catalog extraction for the ALFALFA survey, and especially with the beginning of the next generation large-scale radio surveys, the scope and sophistication of these investigations will grow exponentially. We conclude therefore this thesis with Chapter 6, by highlighting promising directions of future research.

5.1 The velocity width function of ALFALFA galaxies & implications for the nature of dark matter

We use a sample of $10\,744$ ALFALFA galaxies to measure the space density of HI-bearing galaxies as a function of their HI linewidth (velocity width function of galaxies; WF). The measurement extends to widths as low as $w = 20\ \mathrm{km\,s^{-1}}$ (uncorrected for inclination), and finds a "shallow" low-width logarithmic slope for the distribution ($\alpha = -0.85 \pm 0.19$).

We compare the ALFALFA WF with two samples of modeled galaxies, populating high-resolution ΛCDM dark matter simulations. The WFs of the virtual samples diverge from the ALFALFA measurement at widths $w \lesssim 200\ \mathrm{km\,s^{-1}}$, reflecting the "steep" distribution of halo velocities predicted by cold dark matter ($\alpha_{\mathrm{CDM}} \approx -3$). The difference in abundance is a factor of ≈ 8 at $w = 50\ \mathrm{km\,s^{-1}}$ (the resolution limit of the simulations), growing to a factor of ~ 100 when extrapolated to the ALFALFA limit of $w = 20\ \mathrm{km\,s^{-1}}$.

We further consider two solutions to the problem: the first rests on suppressing the formation of low-mass halos, and involves considering alternative dark matter models, such as a keV-scale warm dark matter (WDM) particle. The

second solution is related to the shape of the rotation curves of dwarf galaxies, and does not require a modification of the dark matter model. In particular, the inclination-corrected HI rotational velocity of dwarf galaxies may systematically underestimate the maximum rotational velocity of the host halo. In this case, the ALFALFA WF can be reproduced if the HI-to-halo velocity ratio (v_{HI}/v_{halo}) drops sharply at the low-velocity end. Future tests of this CDM prediction, for example through the measurement of spatially resolved HI kinematics of dwarf galaxies, can yield important clues on the nature of dark matter.

5.2 The baryonic contents of galaxies & constraints on galactic feedback

We use optical data from the Sloan Digital Sky Survey and 21cm data from the ALFALFA survey to measure the "baryonic mass" of galaxies in the local universe ($M_b = M_* + 1.4\,M_{HI}$). We then perform abundance matching between the distributions of stellar, HI and baryonic mass and the halo mass function in the ΛCDM cosmology, to obtain the average relations of M_*/M_h and M_b/M_h as a function of halo mass. The measurement shows that that low-mass halos are significantly baryon depleted, even when their atomic gas component is taken into account. For example, halos with masses $M_h = 10^{10.3}\,M_\odot$ have average baryon fractions that are just 2% of the cosmic value ($\eta_b \approx 0.02$). This result is particularly important in view of the fact that atomic gas often dominates the baryonic mass budget of dwarf galaxies. Moreover, it contrasts with previous estimates, which indicated an approximately constant value of $\eta_b \approx 0.10$ for low halo masses.

Such low values of η_b are difficult to reconcile with current models of galaxy formation. In particular, the halos under consideration are too large to be affected by photoionization heating in the early universe. As a result, additional feedback mechanisms must be in operation, such as star-formation and supernova driven galactic winds. The observed η_b - M_h relation requires extremely efficient feedback, able to expel from the galaxy ~ 100 times more gas than the amount of stars it leaves behind. Such intense feedback is difficult to accommodate within our current understanding of galaxy formation, and correctly reproducing the low baryon fractions of dwarf galaxies may constitute a major challenge for hydrodynamic simulations of galaxy formation.

5.3 The clustering of ALFALFA galaxies & their relationship to optical galaxies and dark matter halos

We use the ALFALFA sample to study the clustering properties of galaxies selected by their atomic hydrogen content (i.e. HI-selected galaxies). We measure the correlation function for subsamples of different HI mass, finding no convincing evidence for variations of clustering with HI mass over the range $M_{HI} = 10^{7.5} - 10^{11}\ M_{\odot}$. Taken at face value, the data indicates that low HI mass galaxies ($M_{HI} = 10^{8.5} - 10^{9.5}\ M_{\odot}$) cluster more weakly than their more HI massive counterparts, but we show that the effect is most probably due to the significantly limited volume probed by these low-mass galaxies ("finite-volume" effect).

We furthermore compare the clustering characteristics of the ALFALFA sample with those of optically selected samples drawn from the SDSS spectroscopic

database. We find that HI-selected galaxies cluster more weakly than optically selected ones, even for absolute magnitudes as low as $M_r \approx -18$. By contrast, when the SDSS sample is restricted to galaxies with blue optical colors, we obtain an extremely good match to the clustering of ALFALFA galaxies. On the other hand, red SDSS galaxies are found to cluster much more strongly: their two-dimensional correlation function shows prominent non-linear features on small scales ("finger of god" effect), which are much more subdued in both the HI and SDSS blue samples. In addition, the cross-correlation function of the ALFALFA and SDSS red samples shows that HI galaxies "avoid" being located in the vicinity of galaxies with red optical colors, and in environments characterized by strong peculiar motions (i.e clusters and rich groups).

Next, we consider the clustering properties of ΛCDM halos selected from the Bolshoi cosmological simulation. There is a clear trend for stronger clustering for more massive halos, unlike in the case of HI galaxies where a similar trend with HI mass is absent. This suggests that HI mass is not tightly related to host halo mass. In addition, we find that the correlation function of halos is incompatible with the ALFALFA correlation function, when subhalos are included in the computation. This result further suggests that a sizable fraction of subhalos do not host gas-rich galaxies. Both findings above are supported by a semi-analytic model of the clustering of HI galaxies, based on abundance matching. In particular, the modeling analysis favors an M_{HI}-M_h relation with large scatter and a weak dependence of galactic HI mass on the mass of the host halo. Lastly, we find that the correlation function of HI-selected galaxies can be reproduced fairly well by just removing low-spin halos from the sample. This finding provides indirect support to the hypothesis that halo spin plays an important role in determining the gas contents of galaxies.

CHAPTER 6

FUTURE WORK

The research presented in this dissertation aims to contribute to our understanding of how galaxies form and evolve, and help elucidate the properties of dark and baryonic matter on galactic scales. For example, some key open questions it aims to answer are:

– What are the properties of dark matter on ∼kpc scales, and what do they entail for the nature of the dark matter particle?
– What are the main physical processes that shape the properties of the present-day galactic population?
– How do we connect observed galaxies with their host dark matter halos?
– How much can we learn about the underlying matter structure of the universe by using galaxies as tracers?

ALFALFA gives us the opportunity to address the questions listed above, through the statistical analysis of the survey's exceptional dataset. Chapters 2, 3 & 4 are examples of this approach, and illustrate how the measurement of the abundance and clustering properties of ALFALFA galaxies can lead to insights into galaxy formation and near-field cosmology issues.

In the future, I plan to take further advantage of the ALFALFA sample and, in combination with other large multiwavelength datasets, use it to advance the research presented in this dissertation. As the ALFALFA sample grows over the next few years –thanks to the release of data covering progressively up to 100% of the survey area– certain investigations will benefit tremendously, especially the ones where sample size is critical. Below is a brief description of the research

topics that I plan to pursue in the short and medium term, and are of relevance to the general issues outlined above.

6.1 The "Baryonic" Tully-Fisher relation of ALFALFA galaxies

The ALFALFA dataset is ideal for the study of the "Baryonic Tully-Fisher" relation (BTFR), which is the observationally established relation between the baryonic mass of galaxies and their rotational velocity. The relation's power-law behavior over several orders of magnitude in mass and its very small observed scatter have been difficult to explain within the ΛCDM paradigm. For example, theoretical work predicts a break down of the power-law at small velocities [e.g. Desmond, 2012, Trujillo-Gomez et al., 2011]; in addition, the distribution of halo concentrations at fixed halo mass is expected to introduce scatter in the BTFR [Dutton, 2012], that can only marginally be accommodated within current observational uncertainties [McGaugh, 2012]. An accurate measurement of the low-velocity behavior of the BTFR and tighter limits on its scatter have therefore great potential for testing our understanding of galaxy formation in the cosmological context. At the same time however, measurements of the BTFR face a number of observational challenges: Firstly, samples of low-mass galaxies with measurements of all three properties that enter the BTFR (stellar mass , gas mass & rotational velocity) are limited in number and heterogeneous in nature. Secondly, measurements of the BTFR are subject to several sources of uncertainty, most importantly systematics related to the conversion of galaxy photometry to stellar mass.

Thanks to its wide sky coverage and superior sensitivity, the ALFALFA survey has the potential to address both limitations. On one hand, the latest pub-

licly available sample already contains >5,000 galaxies with measured 21cm linewidths, as well as stellar and atomic gas masses. The sample is selected in an unbiased and homogeneous fashion, and includes objects with rotational velocities as low as $v_{rot} \approx 20\ \mathrm{km\,s^{-1}}$ and masses as low as $M_{bar} \approx 10^7\,M_\odot$. Furthermore, the ALFALFA dataset includes ~ 100 highly gas-dominated galaxies ($M_{gas} \gtrsim 3\,M_*$), which can be used to provide stringent constraints on the scatter of the BTFR. In particular, gas-dominated galaxies are extremely valuable in BTFR studies because most of their baryonic content is in the form of atomic gas, whose mass can be calculated from their 21cm flux with negligible systematic uncertainty. I therefore plan to use the unique sample of ALFALFA galaxies to study in detail the BTFR, and reduce the observational uncertainties regarding its shape and scatter. The project is currently in its initial stage, but I anticipate delivering concrete results in the first few months of 2014.

6.2 In search of a solution to small-scale CDM challenges: warm dark matter or baryonic physics?

The current "standard" ΛCDM cosmological model $-$involving a cosmological constant (Λ) and non-baryonic dark matter with negligible free streaming length (CDM)$-$ has been remarkably successful at reproducing the large-scale structure of our universe. However, the model is facing a number of important observational challenges on galactic scales: CDM structure formation predicts a multitude of low-mass halos, seemingly in contradiction with the paucity of low-mass galaxies observed in surveys. The interpretation of these observational challenges has been hampered by the fact that the formation and evo-

lution of dwarf galaxies is a complex process, involving the strong interplay between dark matter properties and poorly understood baryonic feedback processes.

The earliest observational challenges to ΛCDM on small scales (e.g. "missing satellites problem", "void phenomenon", etc.) relied on the simplistic assumption of a one-to-one correspondence between dark matter halos and observed galaxies. In recent years however, observational and theoretical work has brought to light the complexity involved in the formation and evolution of galaxies hosted by low-mass halos, and has highlighted the difficulties in comparing observations with theoretical expectations. In response, ΛCDM observational challenges have been evolving towards "cleaner" formulations, which try to separate the effects of baryons and dark matter on the properties of observed galaxies. One of the most important challenges involves the combination of the abundance of low-mass galaxies and their internal kinematics. On one hand, statistical measurements of galaxy abundance require that all CDM halos hosting field galaxies detectable by current wide-area surveys have maximum rotational velocities $v_{halo} > 30\ \mathrm{km\,s^{-1}}$ (see Chapter 2). This "threshold" is necessary because extending galaxy formation to smaller (and much more abundant) halos would result in a dramatic overestimate of the number of visible galaxies. On the other hand, the internal kinematics of a fair fraction of dwarf galaxies indicate that their host halos are well below this mass threshold [Ferrero et al., 2012]. This challenge is the *field analog* of the "too big to fail" problem [Boylan-Kolchin et al., 2012].

One solution to the problem would be suppressing the formation of low-mass halos, which could naturally explain the scarcity of low-mass galaxies and

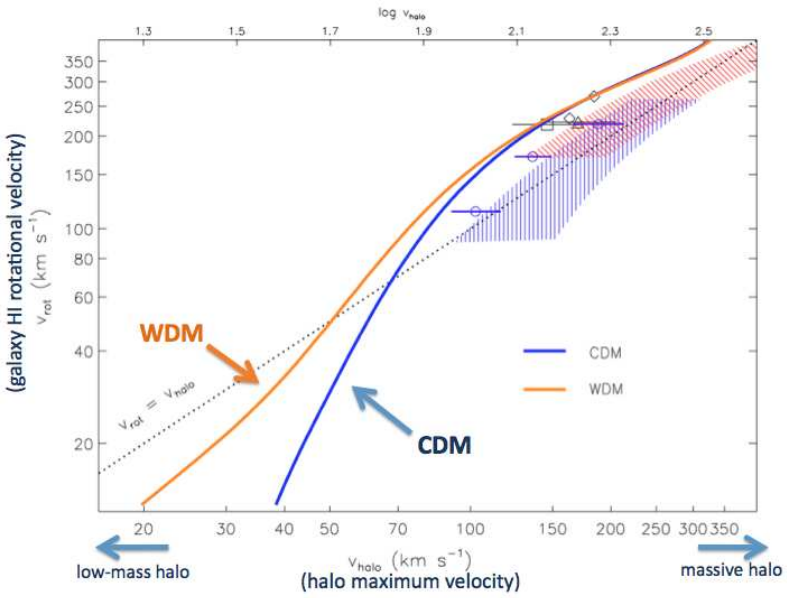

Figure 6.1 The average relation between the maximum rotational velocity of a galaxy as measured by the width of its 21cm emission line (v_{rot}) and the maximum rotational velocity of its host dark matter halo (v_{halo}) in a CDM and WDM universe (blue and orange solid lines, respectively). I plan to test the two models by using resolved HI rotation curves of dwarf galaxies published in the literature, in order to place individual $\{v_{rot}, v_{halo}\}$ datapoints on the crucial low-velocity quadrant of the plot.

would lift the need for a galaxy formation mass "threshold". This is accomplished in alternative dark matter theories, such as warm dark matter (WDM) models with particle masses in the ∼keV range. Figure 6.1 represents a quantitative statement of the issue, and shows that the average relation between the measured HI rotational velocity of a galaxy (v_{rot}) and the host halo mass (v_{halo}) predicted by the CDM and WDM models is very different at low velocities.

I therefore plan to use rotation curves of galaxies, in order to compare the average "$\mathbf{v_{rot}}$ - $\mathbf{v_{halo}}$" relation predicted by the two models to data measured from individual galaxies. In the first phase of the project I plan to use a sample of rotation curves compiled from the literature, which will include the SFI++ database of spiral galaxies [Catinella et al., 2006], as well as ≈ 200 dwarf galaxies with HI kinematics [Côté et al., 2000, McGaugh, 2005, Begum et al., 2008, Oh et al., 2011, Swaters et al., 2009, Trachternach et al., 2009, Wolf et al., 2010, Stark et al., 2009]. I plan to perform mass modeling of the rotation curves of these galaxies in order to determine both their galactic rotational velocity (v_{rot}) and their host halo mass (v_{halo}), in order to place individual datapoints on the v_{rot} - v_{halo} plane. In a subsequent phase of this project, the literature sample mentioned above could be augmented by a sample of ≈30 additional dwarf galaxies of extremely low mass ($v_{rot} \lesssim 30\ \mathrm{km\,s^{-1}}$), selected directly from the ALFALFA survey. Their velocity fields could be measured by using 21cm interferometry and/or wide field of view integral field units on optical telescopes.

6.3 The baryon contents of galaxies and galactic feedback

Baryonic physics have a profound effect on the properties and internal structure of observed galaxies today. In fact, recent hydrodynamical simulations of galaxy formation have proposed that very efficient supernova feedback is the fundamental mechanism setting the scaling of baryonic galaxy properties with host halo mass [e.g. McCarthy et al., 2012, Brook et al., 2012]. However, simulation outcomes depend sensitively on the details of the computational implementation of stellar feedback and galactic outflows. Observational constraints are therefore critical for pinning down the values of key parameters, such as the energy transport efficiency of supernova explosions and the mass of gas swept up by the resulting blasts. Global measurements of the galactic baryon fraction as a function of host halo mass –such as those presented in Chapter 3– are important in this respect, but cannot distinguish between internal feedback processes and environmental effects.

I therefore plan to take advantage of the large optical and 21cm datasets of the SDSS and ALFALFA surveys to determine the dependence of the baryonic contents of galaxies as a function of their environment. A project is already underway to characterize the environment of all ALFALFA galaxies (Michael Jones et al., in prep.). Nearest-neighbor based methods can be used to measure the immediate environment of a galaxy, while statistical methods can measure the density on large scales (e.g. void vs. filament environment, etc.). Once a full catalog of ALFALFA galaxy environments is compiled, the analysis described in Chapter 3 can be extended to address the variation of galactic baryon fractions with environment density. In particular, *isolated* field galaxies are ideal laboratories for studying internal feedback processes, such as supernova-driven gas

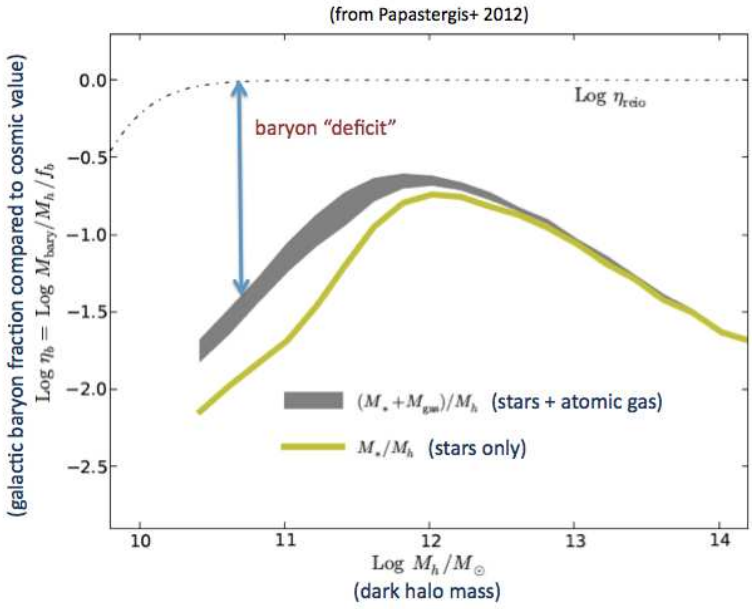

Figure 6.2 The baryon fraction of galaxies (stars + atomic gas) as a function of host halo mass (see Sec. 3.5). Low-mass halos are severely baryon depleted, presumably due to intense galactic outflows. I plan to determine the same distribution for *isolated* low-mass galaxies, delivering stringent constraints on the stellar feedback implementations of hydrodynamical simulations of galaxy formation.

blowout. A measurement of the baryon fraction of these galaxies can shed light on the efficiency of baryonic feedback processes, and inform the numerical implementation of stellar feedback in hydrodynamic simulations of galaxy formation. Conversely, the analysis of samples located in high density environments (e.g. clusters, groups, etc.) can be used to quantify the effect of environment on the baryonic contents of galaxies.

6.4 Advanced modeling of the connection between HI-selected galaxies and their host halos

The detailed modeling of gas-rich galaxies in terms of the population of host halos will be of great value in the upcoming era of large 21cm surveys, both for extragalactic as well as for cosmological studies. Presently, little is known regarding the M_{HI}-M_{halo} relation in the local universe, apart from its average scaling [e.g. Marín et al., 2010, Evoli et al., 2011, Papastergis et al., 2012]. The work presented in Section 4.4.2 already represents a step forward, since it uses the dependence of clustering on HI mass to broadly gauge the scatter of the M_{HI}-M_{halo} relation. Eventually, a full analysis of the connection between HI-galaxies and their host halos should be performed, at a level a detail similar to what is now achieved for optical samples [e.g. Reid et al., 2012, Rodríguez-Puebla et al., 2013]. Such modeling work will benefit greatly from improved observational constraints, including high accuracy measurements of the clustering properties of HI galaxies and measurements of the environmental dependence of the HIMF. Lastly, it is currently largely unknown whether halo properties other than mass (e.g. spin, concentration, merger history, etc.) play an important role in determining the gas contents of galaxies. Section 4.4.3 presents a preliminary investigation of the role of spin in this regard, but the general issue certainly merits a systematic and in-depth investigation.

APPENDIX A

STATISTICAL ESTIMATION OF THE DISTRIBUTION OF GALAXY PROPERTIES

The statistical distributions of galactic properties –such as the luminosity function (LF), stellar, HI and baryonic mass functions (SMF, HIMF & BMF respectively), velocity function (VF), etc.– are invaluable tools for studying the formation and evolution of galaxies and for constraining the properties of dark matter on small scales. An accurate measurement of these distributions, therefore, is of crucial importance for a large number of extragalactic studies. For example, any bias in the measurement of the galaxy abundances presented in Chapters 2, 3 & 4 will directly affect the main results of these works, and could conceivably alter their scientific interpretation. As galaxy sample sizes have grown in size –and counting errors have decreased accordingly–, systematic uncertainties and biases have become the dominant source of error for these type of measurements. As a result, the statistical techniques used to estimate such distributions have progressively become more sophisticated –and less intuitive.

In this Appendix, we present a brief overview of the main techniques used to estimate galactic distributions of luminosity, mass, velocity, etc. We begin in Sec. A.1 by introducing some basic definitions and by presenting the widely-used "$1/V_{max}$" estimator. In Sec A.2, we give an introduction of maximum-likelihood techniques, and we illustrate their setup and implementation in a simple setting. Sec. A.2 also includes the definition of the "$1/V_{eff}$" method, which is used extensively in Chapters 2, 3 & 4 in this dissertation. Laslty, in Sec. A.3 & A.4 we present the specific implementation of the "$1/V_{eff}$" method for the ALFALFA dataset.

Even though the techniques presented below can be used for estimating the distribution of any galactic property (e.g. luminosity, mass, color, etc.) using several types of galactic samples, we consider from now on the specific case of estimating the HIMF from a 21cm selected sample; adapting the methods for use in other settings is relatively straightforward.

A.1 Basic definitions and the "$1/Vmax$" method

The HIMF, denoted by $n(M_{HI})$, is defined as the number density of galaxies with HI masses within an infinitesimal range around M_{HI}. In practice, the HIMF is calculated in bins of HI mass of finite width (either linear or logarithmic), in which case it is expressed as a set of bin values, n_j, where $j = 1, \ldots, N_m$. Calculating the binned HIMF in the case of a volume-limited galaxy sample is extremely straightforward: it just consists of counting the number of sample

galaxies within each HI mass bin (N_j) and normalizing the count by the total survey volume and the bin size, $n_j = N_j/(V_{survey}\,\Delta M_{HI})$.

Most galactic samples, however, are not volume-limited, but rather they are selected based on an apparent flux limit (flux-limited samples). As a result, galaxies with different properties (e.g. HI mass) are detected out to different distances, making the calculation of the HIMF more involved. In the case of a flux-limited sample, the count of sample galaxies within each bin must be normalized by the volume over which galaxies within the specific bin can be detected by the survey, $V_{max,j}$. Therefore, for a flux-limited sample

$$n_j = \frac{N_j}{V_{max,j}\,\Delta M_{HI}} \quad . \tag{A.1}$$

Based on the relation above, Schmidt [1968] developed the "$1/V_{max}$" technique, where the HIMF can be estimated via the expression:

$$n_j = \frac{1}{\Delta M_{HI}} \sum_i \frac{1}{V_{max,i}}, \quad \text{for all galaxies } i \text{ within mass bin } j. \qquad (1/V_{max} \text{ method}) \tag{A.2}$$

In the expression above $V_{max,i}$ is the volume over which galaxy i can be detected by the survey. In the case of a 21cm sample with a flux limit of S_{lim}, the maximum volume for galaxy i is:

$$V_{max,i} = \frac{1}{3}\,\Omega_{survey}\,D^3_{max,i}\ , \tag{A.3}$$

$$\text{where } D_{max,i} = \sqrt{\frac{M_{HI,i}}{k\,S_{lim}}} \qquad (k = 2.356\,10^5\ M_\odot\,(\mathrm{Jy\,kms^{-1}})^{-1}\mathrm{Mpc^{-2}}). \tag{A.4}$$

The Poisson counting error of the $1/V_{max}$ estimates of the bin values is furthermore given by the simple expression:

$$\sigma_j^2 = \frac{1}{\Delta M_{HI}^2} \sum_i \frac{1}{V_{max,i}^2}, \quad \text{for all galaxies } i \text{ within mass bin } j. \tag{A.5}$$

If the sample volume is homogeneous (or nearly homogeneous), the $1/V_{max}$ estimate will converge towards the true HIMF as the size of the sample increases. However, the method suffers from one important disadvantage: the

values of different HIMF bins are determined by counting galaxies detected over different volumes. Therefore, any large-scale structure in the survey volume will affect different bin counts differently, leading to biased estimates of the space densities. This problem is systematic in nature and cannot be reduced by simply increasing the sample size. In order to mitigate the problem, density-independent estimators for the HIMF have been developed, based on maximum likelihood (ML) techniques.

A.2 Maximum-likelihood techniques and the "$1/V_{eff}$" method

Let us denote by $\phi(M_{HI})$ the probability that a galaxy picked from a volume-limited sample has mass M_{HI}. Note that ϕ_{HI} is the same as the HIMF, $n(M_{HI})$, except that the former is normalized to unity. The likelihood that our flux-limited sample contains a galaxy with mass $M_{HI} = M_{HI,i}$ located at a distance D_i is therefore:

$$\ell_i = \frac{\phi(M_{HI,i})}{\int_{M_{HI}=M_{HI,lim}(D_i)}^{\infty} \phi(M_{HI})\, dM_{HI}} \quad . \tag{A.6}$$

The numerator in the expression above is just the volume-limited probability that the detected galaxy has $M_{HI} = M_{HI,i}$, while the integral in the denominator spans the range of M_{HI} that can be detected at distance D_i, given the flux-limited nature of the sample. In particular, $M_{HI,lim}(D_i)$ is given by

$$M_{HI,lim}(D_i) = k\, S_{lim}\, D_i^2 \quad . \tag{A.7}$$

The joint likelihood of detecting each galaxy, $i = 1, \ldots, N_g$, in the sample is then simply the product of the individual likelihoods, $\mathcal{L} = \prod_i \ell_i$. Given an observed galactic sample therefore, $\mathcal{L}$ is just a number that depends on the distribution $\phi(M_{HI})$. If we believe that $\phi(M_{HI})$ follows a certain analytic form, then the value of $\mathcal{L}$ depends just on the parameters of the distribution. For example, the HI mass function has been shown to follow a Schechter functional form within current measurement errors [Martin et al., 2010], in which case:

$$\phi(M_{HI}) = \frac{dP}{dM_{HI}} = \frac{\phi_*}{M_*} \left(\frac{M_{HI}}{M_*}\right)^{\alpha} e^{-\left(\frac{M_{HI}}{M_*}\right)} \quad . \tag{A.8}$$

Equation A.8 approaches a power-law with exponent α for HI masses well below the "characteristic" value M_* (i.e. $\phi \propto M_{HI}^{\alpha}$ for $M_{HI} \ll M_*$), while it ap-

proaches a declining exponential at the high-mass end (i.e. $M_{HI} \propto e^{-\left(\frac{M_{HI}}{M_*}\right)}$ for $M_{HI} \gg M_*$). The parameter ϕ_* sets the overall normalization of the distribution. Under the assumption of a Schechter form, the joint likelihood is therefore a function of just two variables, $\mathcal{L} = \mathcal{L}(M_*, \alpha)$ (note that the normalization parameter ϕ_* drops out of Eqn. A.6). In this instance it is practical to explicitly calculate $\mathcal{L}$ on a grid of values in the $\{M_*, \alpha\}$ plane, and determine the values of the parameters that maximize the joint likelihood (in practice, the maximum of the log-likelihood, $\ln \mathcal{L} = \sum_i \ln \ell_i$, is sought for). The approach outlined above is a *parametric* maximum-likelihood method, because it relies on the assumption that the HIMF follows a specific parametric distribution. Parametric ML methods have been used extensively in the literature to measure a number of galactic distributions, including the LF [Sandage et al., 1979] and the HIMF [Martin et al., 2010].

However, current galaxy sample sizes are large enough that can be used for testing specific functional forms, and potentially detecting small deviations from them. In this case, one would like to use a *non-parametric* approach, where no specific analytic form must be assumed a priori. Such a technique is implemented in Efstathiou et al. [1988], and is referred to as the step-wise maximum likelihood (SWML) method. SWML is based on splitting the HI-mass distribution in bins of log HI-mass ($m = \log(M_{HI}/M_\odot)$), and treating each bin value as a model parameter, $\hat{\phi}_j$ ($j = 1, 2, ..., N_m$). The maximum likelihood estimate for $\phi(M_{HI})$ is therefore obtained by adjusting each $\hat{\phi}_j$ in order to maximize the joint likelihood $\mathcal{L}$.

In this setup, the individual likelihood for each galaxy (Eqn. A.6) is expressed as:

$$\ell_i = \frac{\sum_j \delta_{ij} \hat{\phi}_j}{\sum_j H_{ij} \hat{\phi}_j \Delta m} \quad . \tag{A.9}$$

The set of coefficients δ_{ij} are used to ensure that only the value for the bin to which galaxy i belongs to appears in the numerator, while the set of coefficients H_{ij} is used to enforce the summation to go only over the mass bins that lie above the minimum HI mass detectable at distance D_i. Denoting the HI-mass at the lower boundary of bin j by m_j^- and the HI mass at the upper boundary by m_j^+, the coefficients δ_{ij} and H_{ij} we can be written as:

$$\delta_{ij} = \begin{cases} 1, & \text{if galaxy } i \text{ belongs to mass bin } j \\ 0, & \text{otherwise} \end{cases} \tag{A.10}$$

and

$$H_{ij} = \begin{cases} 0, & m_j^+ < m_{lim,D_i} \\ \frac{m_j^+ - m_{lim,D_i}}{m_j^+ - m_j^-}, & m_j^- < m_{lim,D_i} \leqslant m_j^+ \\ 1, & m_{lim,D_i} \leqslant m_j^- \end{cases} \tag{A.11}$$

The joint log-likelihood can therefore be written as

$$\ln \mathcal{L} = \sum_i \ln \ell_i = \sum_i \sum_j \delta_{ij} \ln(\hat{\phi}_j \Delta m) - \sum_i \ln \left(\sum_j H_{ij} \hat{\phi}_j \Delta m \right) + \text{const.} \tag{A.12}$$

Typically the HIMF is measured in $\sim$tens of bins, such that an explicit calculation of the log-likelihood over a multidimensional grid of $\{\hat{\phi}_j\}$ parameters is no longer practical. Instead, an extremum point is sought for analytically, by setting the partial derivative of the log-likelihood with respect to each parameter equal to zero, $\partial \ln \mathcal{L} / \partial \hat{\phi}_j = 0$, for all j. The solution is then given by (up to an overall normalization factor):

$$\hat{\phi}_j = \frac{\sum_i \delta_{ij}}{\sum_i \frac{H_{ij}}{\sum_m H_{im} \hat{\phi}_m}} = \frac{N_j}{\sum_i \frac{H_{ij}}{\sum_m H_{im} \hat{\phi}_m}} \quad . \tag{A.13}$$

Note that the solution of Eqn. A.13 is expressed in terms of the unknown parameters themselves, so the final solution is found by iterating from an initial guess.

Before proceeding, it is worth giving an intuitive interpretation of Eqn. A.13: Recall that calculating the HIMF involves essentially counting the number of sample galaxies within bin j and normalizing the count by the volume over which these galaxies can be detected. With this in mind, the numerator in Eqn. A.13 should be intuitively clear. This consideration also leads us to the conclusion that the denominator must be playing the role of an "effective" volume. In particular, the denominator consists of a sum over terms that are non-zero only for galaxies for which $H_{ij} \neq 0$. After a close inspection of Eqn. A.11 we can see that $H_{ij} \neq 0$ only for sample galaxies that are located within the volume over which galaxies belonging to mass bin j can be detected. Furthermore, each galaxy count is weighted by $1/\sum_m H_{im} \hat{\phi}_m$. This weight is the inverse of the survey selection function (see Sec. 4.2.3), i.e. it is used to take into account the fact

that each detection at distance D is accompanied by a number of $1/\sum_m H_{im}\hat{\phi}_m$ fainter galaxies at this distance that are not detected. In essence therefore, the denominator is the number of galaxies (corrected for the survey selection function) residing in the volume accessible to galaxies with masses within bin j. In the case of a homogeneous sample, it is easy to see why this quantity acts effectively as a volume: the number of objects calculated in this way is simply proportional to the physical volume over which galaxies in bin j are detectable. In this case, the SWML method is equivalent to the $1/V_{max}$ method, apart from an overall normalization. However, the SWML estimate is less prone to inhomogeneities in the survey volume, because any density fluctuations will affect the numerator and denominator in Eqn. A.13 in the same way, therefore preserving their ratio.

Taking advantage of this intuitive picture, Zwaan et al. [2005] developed the "$1/V_{eff}$" ML method. In particular, the parameters $\hat{\phi}_j$ can be obtained (up to an overall normalization) by the expression

$$\hat{\phi}_j = \sum_i \frac{1}{V_{eff,i}}, \quad \text{for all galaxies } i \text{ within mass bin } j, \qquad (1/V_{eff} \text{ method}) \tag{A.14}$$

where

$$V_{eff,i} = \sum_k \frac{1}{\sum_m H_{km}\hat{\phi}_m}, \tag{A.15}$$

for all galaxies k located within the volume over which galaxy i is detectable. Accordingly, the counting error in the $1/V_{eff}$ method is given by:

$$\sigma_j^2 = \sum_i \frac{1}{V_{eff,i}^2}, \quad \text{for all galaxies } i \text{ within mass bin } j. \qquad (1/V_{eff} \text{ error}) \tag{A.16}$$

Note that the similarity of Eqns. A.14 & A.16 with Eqns. A.2 & A.5 is the reason for the method's "$1/V_{eff}$" name.

A.3 The bivariate case and implementation for the ALFALFA dataset

In any blind 21cm survey (i.e. any survey that spends an approximately constant amount of time on each pointing), the detectability of an HI source depends not only on its integrated flux, S_{int}, but also on the width of its line profile, W_{50}. This is because for a fixed value of S_{int}, sources with larger W_{50} will have their spectral profile affected by noise over a larger bandwidth. As a result, the overall noise level affecting a detection is expected to scale with source velocity width as $\propto W_{50}^{1/2}$, and the corresponding signal-to-noise ratio is expected to scale as $S/N \propto S_{int}\, W_{50}^{-1/2}$.

As a result, the ALFALFA sample is not a purely flux-limited one, but sources are included in the sample based on their combination of integrated flux and velocity width. The completeness of the ALFALFA survey on the $\{S_{int}, W_{50}\}$-plane was quantitatively determined by Haynes et al. [2011, Sec. 6], based on the observed sample itself. In actuality, the ALFALFA completeness is not an idealized step function in S_{int}, but rather a smooth surface, $C(S_{int}, W_{50})$, which Haynes et al. [2011] characterize by giving the 90%, 50% and 25% completeness contours (see their Eqns. 4 & 5). Under these circumstances the procedure outlined in Sec. A.2 has to be generalized to a bivariate maximum-likelihood method. In particular, the detection likelihood for individual galaxies stemming from Eqn. A.6 now becomes:

$$\ell_i = \frac{\phi(M_{HI,i}, W_{50,i})}{\int_{W_{50}=0}^{\infty} \int_{M_{HI}=0}^{\infty} C_i(M_{HI}, W_{50})\, \phi(M_{HI}, W_{50})\, dM_{HI} dW_{50}} \quad . \tag{A.17}$$

In the equation above, $C_i(M_{HI}, W_{50})$ is the ALFALFA completeness surface $C(S_{int}, W_{50})$ expressed in terms of the HI mass at the distance of galaxy i, through the relation $M_{HI} = k\, S_{int}\, D_i^2$.

We proceed by splitting $\phi(M_{HI}, W_{50})$ in logarithmic bins of mass and width, $m = \log(M_{HI}/M_\odot)$ and $w = \log(W_{50}/\mathrm{km\,s^{-1}})$, assuming a constant value within each bin. This leads to the two-dimensional step wise maximum likelihood (2DSWML) technique, where the parameters of the two-dimensional distribution can now be written as $\hat{\phi}_{jk}$ ($j = 1, 2, ..., N_m$ and $k = 1, 2, ..., N_w$). The individual likelihood for each galaxy (analogous to Eqn. A.9) becomes

$$\ell_i = \frac{\sum_j \sum_k \delta_{ijk} \hat{\phi}_{jk}}{\sum_j \sum_k H_{ijk} \hat{\phi}_{jk} \Delta m \Delta w} \quad , \tag{A.18}$$

where the set of coefficients δ_{ijk} and H_{ijk} are defined by analogy to Eqns. A.10 & A.11:

$$\delta_{ijk} = \begin{cases} 1, & \text{if galaxy } i \text{ belongs to mass bin } j \text{ and width bin } k \\ 0, & \text{otherwise} \end{cases} \tag{A.19}$$

and

$$H_{ijk} = \frac{1}{\Delta m \Delta w} \int_{w_k^-}^{w_k^+} \int_{m_j^-}^{m_j^+} C_i(m, w)\, dm dw \quad , \tag{A.20}$$

where m_j^- and m_j^+ are the HI mass at the lower and upper boundary of mass bin j correspondingly and similarly w_k^- and w_k^+ are the upper and lower boundaries of width bin k.

In the chapters of this dissertation we approximate the full $C_i(m, w)$ surface with a step function at the 50% completeness contour level of ALFALFA, therefore excluding all galaxies detected at less than 50% completeness from the computation. Rosenberg and Schneider [2002] find that this approximation reproduces well the HIMF obtained by considering the full shape of the completeness surface [but see Obreschkow et al., 2013, for potential issues when working with large samples].

Due to the close analogy of Eqn A.18 with Eqn. A.9, the remainder of the derivation follows closely the flow described in Sec. A.2. In particular, the joint log-likelihood can now be expressed as:

$$\begin{aligned} \ln \mathcal{L} = \sum_i \ln \ell_i = \sum_i \sum_j \sum_k \delta_{ijk} \ln(\hat{\phi}_{jk} \Delta m \Delta w) \\ - \sum_i \ln \left(\sum_j \sum_k H_{ijk} \hat{\phi}_{jk} \Delta m \Delta w \right) + \text{const.} \end{aligned} \tag{A.21}$$

$\ln \mathcal{L}$ is maximized by setting the partial derivatives with respect to each of the parameters, $\hat{\phi}_{jk}$, equal to zero. This leads to the relation:

$$\hat{\phi}_{jk} = \frac{\sum_i \delta_{ijk}}{\sum_i \frac{H_{ijk}}{\sum_m \sum_n H_{imn} \hat{\phi}_{mn}}} = \frac{N_{jk}}{\sum_i \frac{H_{ijk}}{\sum_m \sum_n H_{imn} \hat{\phi}_{mn}}} \quad . \tag{A.22}$$

By analogy to Eqn. A.13, the denominator in the expression above corresponds to the "effective" volume available to galaxies belonging to mass bin j and width bin k. We can therefore generalize the $1/V_{eff}$ method described in Sec. A.2 to take into account of the bivariate mass-width distribution of galaxies in the following way:

$$\hat{\phi}_{jk} = \sum_{i} \frac{1}{V_{eff,i}}, \quad \text{for all galaxies } i \text{ within mass bin } j \text{ and width bin } k, \tag{A.23}$$

where the effective volume for galaxy i is now given by

$$V_{eff,i} = \sum_{l} \frac{1}{\sum_m \sum_n H_{lmn} \hat{\phi}_{mn}}, \tag{A.24}$$

for all galaxies l located within the volume over which galaxy i is detectable. The counting error is given (as per Eqn. A.16) by

$$\sigma^2_{jk} = \sum_{i} \frac{1}{V^2_{eff,i}}, \quad \text{for all galaxies } i \text{ within mass bin } j \text{ and width bin } k. \tag{A.25}$$

A.4 Recovering the HIMF Normalization

As Eqns. A.6 & A.17 imply, the overall normalization of the HIMF cannot be determined through the ML methods described above. As a result, only the relative values of the $\hat{\phi}$ parameters calculated by ML estimators are meaningful. In order to obtain values of the binned HIMF with units of density, we re-scale the $\hat{\phi}_{jk}$ parameters by a common factor, $n_{jk} = \alpha \hat{\phi}_{jk}$, such that the integral of the mass-width distribution matches the inferred average number density of galaxies in the survey volume, $\bar{n}$.

$$\sum_{j} \sum_{k} n_{jk} \, \Delta m \Delta w = \bar{n} \quad . \tag{A.26}$$

Davis and Huchra [1982] discuss various estimators for $\bar{n}$ that strike different balances between stability against poor knowledge of the selection function of the survey, $S(D)$, and immunity to large-scale structure. In this work we choose

to adopt the estimator that is least prone to bias, denoted by n_1 in Davis and Huchra [1982], defined as

$$\bar{n} = n_1 = V_{survey}^{-1} \int \frac{N(D)\, dD}{S(D)} \quad . \tag{A.27}$$

In expression A.27, $N(D)\, dD$ is the number of sample galaxies detected in a spherical shell of thickness dD and radius D, and V_{survey} is the total survey volume. The selection function $S(D)$ is the fraction of galaxies detectable at distance D and is given by

$$S(D) = \frac{\int_{w_{min}}^{w_{max}} \int_{m_{lim}(w,D)}^{m_{max}} \phi(m,w)\, dm\, dw}{\int\int \phi(m,w)\, dm\, dw} \quad , \tag{A.28}$$

where $m_{lim}(w, D)$ is the minimum detectable HI mass[1] for a source with velocity width w at distance D.

In the bivariate step-wise case, which is relevant in the context of the ALFALFA sample, n_1 is evaluated through the expression

$$n_1 = V_{survey}^{-1} \sum_{i=1}^{N_g} \frac{1}{\sum_j \sum_k H_{ijk} \hat{\phi}_{jk} \Delta m \Delta w} \quad , \tag{A.29}$$

provided that the $\hat{\phi}_{jk}$ are normalized to unity, $\sum_j \sum_k \hat{\phi}_{jk}\, \Delta m \Delta w = 1$. In essence, Eqn. A.29 corresponds to weighing each detected galaxy in the survey volume by the inverse of the selection function at the galaxy's distance, effectively correcting each detection by the fraction of galaxies that cannot be detected at distance D_i.

Lastly, we calculate the HIMF from the properly scaled set of n_{jk} values (according to Eqn. A.26) by marginalizing over velocity width:

$$n_j = \sum_{k=1}^{N_w} n_{jk}\, \Delta w \quad , \tag{A.30}$$

[1]The minimum HI mass value, $m_{lim}(w, D)$, can only be defined when the full ALFALFA completeness surface is approximated as a step-function at some completeness level. In the general case, the numerator of Eqn. A.28 is $\int_{w_{min}}^{w_{max}} \int_{m_{min}}^{m_{max}} C_D(m,w)\, \phi(m,w)\, dm\, dw$, where $C_D(m,w)$ is the completeness function in the mass-width plane at distance D.

while the counting error is similarly obtained as

$$\sigma_j^2 = \sum_{k=1}^{N_w} \sigma_{jk}^2 \, \Delta w^2 \quad . \tag{A.31}$$

*The appendix has partial overlap with Appendix B of the published article Martin et al. (2010).

BIBLIOGRAPHY

K. Abazajian and S. M. Koushiappas. Constraints on sterile neutrino dark matter. *Phys. Rev. D*, 74(2):023527–+, July 2006. doi: 10.1103/PhysRevD.74.023527.

K. N. Abazajian, J. K. Adelman-McCarthy, M. A. Agüeros, S. S. Allam, C. Allende Prieto, D. An, K. S. J. Anderson, S. F. Anderson, J. Annis, N. A. Bahcall, and et al. The Seventh Data Release of the Sloan Digital Sky Survey. *ApJS*, 182:543, June 2009. doi: 10.1088/0067-0049/182/2/543.

A. A. Abdo, M. Ackermann, M. Ajello, W. B. Atwood, L. Baldini, J. Ballet, G. Barbiellini, D. Bastieri, K. Bechtol, R. Bellazzini, B. Berenji, E. D. Bloom, E. Bonamente, A. W. Borgland, A. Bouvier, J. Bregeon, A. Brez, M. Brigida, P. Bruel, T. H. Burnett, S. Buson, G. A. Caliandro, R. A. Cameron, P. A. Caraveo, S. Carrigan, J. M. Casandjian, C. Cecchi, Ö. Çelik, A. Chekhtman, J. Chiang, S. Ciprini, R. Claus, J. Cohen-Tanugi, J. Conrad, C. D. Dermer, A. de Angelis, F. de Palma, S. W. Digel, E. Do Couto E Silva, P. S. Drell, A. Drlica-Wagner, R. Dubois, D. Dumora, Y. Edmonds, R. Essig, C. Farnier, C. Favuzzi, S. J. Fegan, W. B. Focke, P. Fortin, M. Frailis, Y. Fukazawa, S. Funk, P. Fusco, F. Gargano, D. Gasparrini, N. Gehrels, S. Germani, N. Giglietto, F. Giordano, T. Glanzman, G. Godfrey, I. A. Grenier, J. E. Grove, L. Guillemot, S. Guiriec, M. Gustafsson, D. Hadasch, A. K. Harding, D. Horan, R. E. Hughes, M. S. Jackson, G. Jóhannesson, A. S. Johnson, R. P. Johnson, W. N. Johnson, T. Kamae, H. Katagiri, J. Kataoka, N. Kawai, M. Kerr, J. Knödlseder, M. Kuss, J. Lande, L. Latronico, M. Llena Garde, F. Longo, F. Loparco, B. Lott, M. N. Lovellette, P. Lubrano, A. Makeev, M. N. Mazziotta, J. E. McEnery, C. Meurer, P. F. Michelson, W. Mitthumsiri, T. Mizuno, A. A. Moiseev, C. Monte, M. E. Monzani, A. Morselli, I. V. Moskalenko, S. Murgia, P. L. Nolan, J. P. Norris, E. Nuss, T. Ohsugi, N. Omodei, E. Orlando, J. F. Ormes, M. Ozaki, D. Paneque, J. H. Panetta, D. Parent, V. Pelassa, M. Pepe, M. Pesce-Rollins, F. Piron, S. Rainò, R. Rando, M. Razzano, A. Reimer, O. Reimer, T. Reposeur, J. Ripken, S. Ritz, A. Y. Rodriguez, M. Roth, H. F.-W. Sadrozinski, A. Sander, P. M. S. Parkinson, J. D. Scargle, T. L. Schalk, A. Sellerholm, C. Sgrò, E. J. Siskind, D. A. Smith, P. D. Smith, G. Spandre, P. Spinelli, J.-L. Starck, M. S. Strickman, D. J. Suson, H. Tajima, H. Takahashi, T. Tanaka, J. B. Thayer, J. G. Thayer, L. Tibaldo, D. F. Torres, Y. Uchiyama, T. L. Usher, V. Vasileiou, N. Vilchez, V. Vitale, A. P. Waite, P. Wang, B. L. Winer, K. S. Wood, T. Ylinen, and M. Ziegler. Fermi Large Area Telescope Search for Photon Lines from 30 to 200 GeV and Dark Matter Implications. *Physical Review Letters*, 104(9): 091302–+, March 2010a. doi: 10.1103/PhysRevLett.104.091302.

A. A. Abdo, M. Ackermann, M. Ajello, W. B. Atwood, L. Baldini, J. Ballet, G. Barbiellini, D. Bastieri, K. Bechtol, R. Bellazzini, B. Berenji, E. D. Bloom, E. Bonamente, A. W. Borgland, J. Bregeon, A. Brez, M. Brigida, P. Bruel, T. H. Burnett, S. Buson, G. A. Caliandro, R. A. Cameron, P. A. Caraveo, J. M. Casandjian, C. Cecchi, A. Chekhtman, C. C. Cheung, J. Chiang, S. Ciprini, R. Claus,

J. Cohen-Tanugi, J. Conrad, A. de Angelis, F. de Palma, S. W. Digel, E. d. C. e. Silva, P. S. Drell, A. Drlica-Wagner, R. Dubois, D. Dumora, C. Farnier, C. Favuzzi, S. J. Fegan, W. B. Focke, P. Fortin, M. Frailis, Y. Fukazawa, P. Fusco, F. Gargano, N. Gehrels, S. Germani, B. Giebels, N. Giglietto, F. Giordano, T. Glanzman, G. Godfrey, I. A. Grenier, J. E. Grove, L. Guillemot, S. Guiriec, M. Gustafsson, A. K. Harding, E. Hays, D. Horan, R. E. Hughes, M. S. Jackson, T. E. Jeltema, G. Jóhannesson, A. S. Johnson, R. P. Johnson, W. N. Johnson, T. Kamae, H. Katagiri, J. Kataoka, M. Kerr, J. Knödlseder, M. Kuss, J. Lande, L. Latronico, M. Lemoine-Goumard, F. Longo, F. Loparco, B. Lott, M. N. Lovellette, P. Lubrano, G. M. Madejski, A. Makeev, M. N. Mazziotta, J. E. McEnery, C. Meurer, P. F. Michelson, W. Mitthumsiri, T. Mizuno, A. A. Moiseev, C. Monte, M. E. Monzani, E. Moretti, A. Morselli, I. V. Moskalenko, S. Murgia, P. L. Nolan, J. P. Norris, E. Nuss, T. Ohsugi, N. Omodei, E. Orlando, J. F. Ormes, D. Paneque, J. H. Panetta, D. Parent, V. Pelassa, M. Pepe, M. Pesce-Rollins, F. Piron, T. A. Porter, S. Profumo, S. Rainò, R. Rando, M. Razzano, A. Reimer, O. Reimer, T. Reposeur, S. Ritz, A. Y. Rodriguez, M. Roth, H. F.-W. Sadrozinski, A. Sander, P. M. Saz Parkinson, J. D. Scargle, T. L. Schalk, A. Sellerholm, C. Sgrò, E. J. Siskind, D. A. Smith, P. D. Smith, G. Spandre, P. Spinelli, M. S. Strickman, D. J. Suson, H. Takahashi, T. Takahashi, T. Tanaka, J. B. Thayer, J. G. Thayer, D. J. Thompson, L. Tibaldo, D. F. Torres, A. Tramacere, Y. Uchiyama, T. L. Usher, V. Vasileiou, N. Vilchez, V. Vitale, A. P. Waite, P. Wang, B. L. Winer, K. S. Wood, T. Ylinen, M. Ziegler, J. S. Bullock, M. Kaplinghat, and G. D. Martinez. Observations of Milky Way Dwarf Spheroidal Galaxies with the Fermi-Large Area Telescope Detector and Constraints on Dark Matter Models. *ApJ*, 712:147–158, March 2010b. doi: 10.1088/0004-637X/712/1/147.

O. Adriani, G. C. Barbarino, G. A. Bazilevskaya, R. Bellotti, M. Boezio, E. A. Bogomolov, L. Bonechi, M. Bongi, V. Bonvicini, S. Bottai, A. Bruno, F. Cafagna, D. Campana, P. Carlson, M. Casolino, G. Castellini, M. P. de Pascale, G. de Rosa, N. de Simone, V. di Felice, A. M. Galper, L. Grishantseva, P. Hofverberg, S. V. Koldashov, S. Y. Krutkov, A. N. Kvashnin, A. Leonov, V. Malvezzi, L. Marcelli, W. Menn, V. V. Mikhailov, E. Mocchiutti, S. Orsi, G. Osteria, P. Papini, M. Pearce, P. Picozza, M. Ricci, S. B. Ricciarini, M. Simon, R. Sparvoli, P. Spillantini, Y. I. Stozhkov, A. Vacchi, E. Vannuccini, G. Vasilyev, S. A. Voronov, Y. T. Yurkin, G. Zampa, N. Zampa, and V. G. Zverev. An anomalous positron abundance in cosmic rays with energies 1.5-100GeV. *Nature*, 458: 607–609, April 2009. doi: 10.1038/nature07942.

Z. Ahmed, D. S. Akerib, S. Arrenberg, M. J. Attisha, C. N. Bailey, L. Baudis, D. A. Bauer, J. Beaty, P. L. Brink, T. Bruch, R. Bunker, S. Burke, B. Cabrera, D. O. Caldwell, J. Cooley, P. Cushman, F. Dejongh, M. R. Dragowsky, L. Duong, J. Emes, E. Figueroa-Feliciano, J. Filippini, M. Fritts, R. J. Gaitskell, S. R. Golwala, D. R. Grant, J. Hall, R. Hennings-Yeomans, S. Hertel, D. Holmgren, M. E.

Huber, R. Mahapatra, V. Mandic, K. A. McCarthy, N. Mirabolfathi, H. Nelson, L. Novak, R. W. Ogburn, M. Pyle, X. Qiu, E. Ramberg, W. Rau, A. Reisetter, T. Saab, B. Sadoulet, J. Sander, R. Schmitt, R. W. Schnee, D. N. Seitz, B. Serfass, A. Sirois, K. M. Sundqvist, M. Tarka, A. Tomada, G. Wang, S. Yellin, J. Yoo, and B. A. Young. Search for Weakly Interacting Massive Particles with the First Five-Tower Data from the Cryogenic Dark Matter Search at the Soudan Underground Laboratory. *Physical Review Letters*, 102(1):011301–+, January 2009. doi: 10.1103/PhysRevLett.102.011301.

L. Anderson, E. Aubourg, S. Bailey, D. Bizyaev, M. Blanton, A. S. Bolton, J. Brinkmann, J. R. Brownstein, A. Burden, A. J. Cuesta, L. A. N. da Costa, K. S. Dawson, R. de Putter, D. J. Eisenstein, J. E. Gunn, H. Guo, J.-C. Hamilton, P. Harding, S. Ho, K. Honscheid, E. Kazin, D. Kirkby, J.-P. Kneib, A. Labatie, C. Loomis, R. H. Lupton, E. Malanushenko, V. Malanushenko, R. Mandelbaum, M. Manera, C. Maraston, C. K. McBride, K. T. Mehta, O. Mena, F. Montesano, D. Muna, R. C. Nichol, S. E. Nuza, M. D. Olmstead, D. Oravetz, N. Padmanabhan, N. Palanque-Delabrouille, K. Pan, J. Parejko, I. Pâris, W. J. Percival, P. Petitjean, F. Prada, B. Reid, N. A. Roe, A. J. Ross, N. P. Ross, L. Samushia, A. G. Sánchez, D. J. Schlegel, D. P. Schneider, C. G. Scóccola, H.-J. Seo, E. S. Sheldon, A. Simmons, R. A. Skibba, M. A. Strauss, M. E. C. Swanson, D. Thomas, J. L. Tinker, R. Tojeiro, M. V. Magaña, L. Verde, C. Wagner, D. A. Wake, B. A. Weaver, D. H. Weinberg, M. White, X. Xu, C. Yèche, I. Zehavi, and G.-B. Zhao. The clustering of galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey: baryon acoustic oscillations in the Data Release 9 spectroscopic galaxy sample. *MNRAS*, 427:3435–3467, December 2012. doi: 10.1111/j.1365-2966.2012.22066.x.

J. Angle, E. Aprile, F. Arneodo, L. Baudis, A. Bernstein, A. Bolozdynya, P. Brusov, L. C. C. Coelho, C. E. Dahl, L. Deviveiros, A. D. Ferella, L. M. P. Fernandes, S. Fiorucci, R. J. Gaitskell, K. L. Giboni, R. Gomez, R. Hasty, L. Kastens, J. Kwong, J. A. M. Lopes, N. Madden, A. Manalaysay, A. Manzur, D. N. McKinsey, M. E. Monzani, K. Ni, U. Oberlack, J. Orboeck, G. Plante, R. Santorelli, J. M. F. Dos Santos, P. Shagin, T. Shutt, P. Sorensen, S. Schulte, C. Winant, and M. Yamashita. First Results from the XENON10 Dark Matter Experiment at the Gran Sasso National Laboratory. *Physical Review Letters*, 100(2):021303–+, January 2008a. doi: 10.1103/PhysRevLett.100.021303.

J. Angle, E. Aprile, F. Arneodo, L. Baudis, A. Bernstein, A. Bolozdynya, L. C. C. Coelho, C. E. Dahl, L. Deviveiros, A. D. Ferella, L. M. P. Fernandes, S. Fiorucci, R. J. Gaitskell, K. L. Giboni, R. Gomez, R. Hasty, L. Kastens, J. Kwong, J. A. M. Lopes, N. Madden, A. Manalaysay, A. Manzur, D. N. McKinsey, M. E. Monzani, K. Ni, U. Oberlack, J. Orboeck, G. Plante, R. Santorelli, J. M. F. Dos Santos, P. Shagin, T. Shutt, P. Sorensen, S. Schulte, C. Winant, and M. Yamashita. Limits on Spin-Dependent WIMP-Nucleon Cross Sections from

the XENON10 Experiment. *Physical Review Letters*, 101(9):091301–+, August 2008b. doi: 10.1103/PhysRevLett.101.091301.

T. Asaka, S. Blanchet, and M. Shaposhnikov. Editors pp CO2 The @nMSM, dark matter and neutrino masses [rapid communication]. *Physics Letters B*, 631:151–156, December 2005. doi: 10.1016/j.physletb.2005.09.070.

D. Aubert, C. Pichon, and S. Colombi. The origin and implications of dark matter anisotropic cosmic infall on $\sim L_*$ haloes. *MNRAS*, 352:376–398, August 2004. doi: 10.1111/j.1365-2966.2004.07883.x.

I. K. Baldry, K. Glazebrook, and S. P. Driver. On the galaxy stellar mass function, the mass-metallicity relation and the implied baryonic mass function. *MNRAS*, 388:945–959, August 2008. doi: 10.1111/j.1365-2966.2008.13348.x.

I. K. Baldry, S. P. Driver, J. Loveday, E. N. Taylor, L. S. Kelvin, J. Liske, P. Norberg, A. S. G. Robotham, S. Brough, A. M. Hopkins, S. P. Bamford, J. A. Peacock, J. Bland-Hawthorn, C. J. Conselice, S. M. Croom, D. H. Jones, H. R. Parkinson, C. C. Popescu, M. Prescott, R. G. Sharp, and R. J. Tuffs. Galaxy And Mass Assembly (GAMA): the galaxy stellar mass function at $z < 0.06$. *MNRAS*, 421:621–634, March 2012. doi: 10.1111/j.1365-2966.2012.20340.x.

S. Basilakos, M. Plionis, K. Kovač, and N. Voglis. Large-scale structure in the HI Parkes All-Sky Survey: filling the voids with HI galaxies? *MNRAS*, 378: 301–308, June 2007. doi: 10.1111/j.1365-2966.2007.11781.x.

A. Begum, J. N. Chengalur, I. D. Karachentsev, and M. E. Sharina. Baryonic Tully-Fisher relation for extremely low mass Galaxies. *MNRAS*, 386:138–144, May 2008. doi: 10.1111/j.1365-2966.2008.13010.x.

P. S. Behroozi, C. Conroy, and R. H. Wechsler. A Comprehensive Analysis of Uncertainties Affecting the Stellar Mass-Halo Mass Relation for $0 < z < 4$. *ApJ*, 717:379–403, July 2010. doi: 10.1088/0004-637X/717/1/379.

P. S. Behroozi, R. H. Wechsler, and H.-Y. Wu. The ROCKSTAR Phase-space Temporal Halo Finder and the Velocity Offsets of Cluster Cores. *ApJ*, 762:109, January 2013a. doi: 10.1088/0004-637X/762/2/109.

P. S. Behroozi, R. H. Wechsler, H.-Y. Wu, M. T. Busha, A. A. Klypin, and J. R. Primack. Gravitationally Consistent Halo Catalogs and Merger Trees for Precision Cosmology. *ApJ*, 763:18, January 2013b. doi: 10.1088/0004-637X/763/1/18.

E. F. Bell, D. H. McIntosh, N. Katz, and M. D. Weinberg. The Optical and Near-Infrared Properties of Galaxies. I. Luminosity and Stellar Mass Functions. *ApJS*, 149:289–312, December 2003. doi: 10.1086/378847.

A. J. Benson and R. Bower. Galaxy formation spanning cosmic history. *MNRAS*, 405:1573–1623, July 2010. doi: 10.1111/j.1365-2966.2010.16592.x.

F. Beutler, C. Blake, M. Colless, D. H. Jones, L. Staveley-Smith, L. Campbell, Q. Parker, W. Saunders, and F. Watson. The 6dF Galaxy Survey: baryon acoustic oscillations and the local Hubble constant. *MNRAS*, 416:3017–3032, October 2011. doi: 10.1111/j.1365-2966.2011.19250.x.

F. Beutler, C. Blake, M. Colless, D. H. Jones, L. Staveley-Smith, L. Campbell, Q. Parker, W. Saunders, and F. Watson. The 6dF Galaxy Survey: dependence of halo occupation on stellar mass. *MNRAS*, 429:3604–3618, March 2013. doi: 10.1093/mnras/sts637.

C. Blake, E. A. Kazin, F. Beutler, T. M. Davis, D. Parkinson, S. Brough, M. Colless, C. Contreras, W. Couch, S. Croom, D. Croton, M. J. Drinkwater, K. Forster, D. Gilbank, M. Gladders, K. Glazebrook, B. Jelliffe, R. J. Jurek, I.-H. Li, B. Madore, D. C. Martin, K. Pimbblet, G. B. Poole, M. Pracy, R. Sharp, E. Wisnioski, D. Woods, T. K. Wyder, and H. K. C. Yee. The WiggleZ Dark Energy Survey: mapping the distance-redshift relation with baryon acoustic oscillations. *MNRAS*, 418:1707–1724, December 2011. doi: 10.1111/j.1365-2966.2011.19592.x.

M. R. Blanton, R. H. Lupton, D. J. Schlegel, M. A. Strauss, J. Brinkmann, M. Fukugita, and J. Loveday. The Properties and Luminosity Function of Extremely Low Luminosity Galaxies. *ApJ*, 631:208–230, September 2005a. doi: 10.1086/431416.

M. R. Blanton, D. J. Schlegel, M. A. Strauss, J. Brinkmann, D. Finkbeiner, M. Fukugita, J. E. Gunn, D. W. Hogg, Ž. Ivezić, G. R. Knapp, R. H. Lupton, J. A. Munn, D. P. Schneider, M. Tegmark, and I. Zehavi. New York University Value-Added Galaxy Catalog: A Galaxy Catalog Based on New Public Surveys. *AJ*, 129:2562–2578, June 2005b. doi: 10.1086/429803.

S. Boissier and N. Prantzos. Chemo-spectrophotometric evolution of spiral galaxies - II. Main properties of present-day disc galaxies. *MNRAS*, 312:398–416, February 2000. doi: 10.1046/j.1365-8711.2000.03133.x.

A. Boselli, J. Lequeux, and G. Gavazzi. Molecular gas in normal late-type galaxies. *A&A*, 384:33–47, March 2002. doi: 10.1051/0004-6361:20011747.

R. G. Bower, A. J. Benson, R. Malbon, J. C. Helly, C. S. Frenk, C. M. Baugh, S. Cole, and C. G. Lacey. Breaking the hierarchy of galaxy formation. *MNRAS*, 370:645–655, August 2006. doi: 10.1111/j.1365-2966.2006.10519.x.

A. Boyarsky, J. Lesgourgues, O. Ruchayskiy, and M. Viel. Realistic Sterile Neutrino Dark Matter with KeV Mass does not Contradict Cosmological Bounds.

Physical Review Letters, 102(20):201304–+, May 2009. doi: 10.1103/PhysRevLett.102.201304.

M. Boylan-Kolchin, V. Springel, S. D. M. White, A. Jenkins, and G. Lemson. Resolving cosmic structure formation with the Millennium-II Simulation. *MNRAS*, 398:1150–1164, September 2009. doi: 10.1111/j.1365-2966.2009.15191.x.

M. Boylan-Kolchin, J. S. Bullock, and M. Kaplinghat. The Milky Way's bright satellites as an apparent failure of ΛCDM. *MNRAS*, 422:1203–1218, May 2012. doi: 10.1111/j.1365-2966.2012.20695.x.

J. Brinchmann, S. Charlot, S. D. M. White, C. Tremonti, G. Kauffmann, T. Heckman, and J. Brinkmann. The physical properties of star-forming galaxies in the low-redshift Universe. *MNRAS*, 351:1151–1179, July 2004. doi: 10.1111/j.1365-2966.2004.07881.x.

C. B. Brook, G. Stinson, B. K. Gibson, J. Wadsley, and T. Quinn. MaGICC discs: matching observed galaxy relationships over a wide stellar mass range. *MNRAS*, 424:1275–1283, August 2012. doi: 10.1111/j.1365-2966.2012.21306.x.

G. Bruzual and S. Charlot. Stellar population synthesis at the resolution of 2003. *MNRAS*, 344:1000–1028, October 2003. doi: 10.1046/j.1365-8711.2003.06897.x.

N. G. Busca, T. Delubac, J. Rich, S. Bailey, A. Font-Ribera, D. Kirkby, J.-M. Le Goff, M. M. Pieri, A. Slosar, É. Aubourg, J. E. Bautista, D. Bizyaev, M. Blomqvist, A. S. Bolton, J. Bovy, H. Brewington, A. Borde, J. Brinkmann, B. Carithers, R. A. C. Croft, K. S. Dawson, G. Ebelke, D. J. Eisenstein, J.-C. Hamilton, S. Ho, D. W. Hogg, K. Honscheid, K.-G. Lee, B. Lundgren, E. Malanushenko, V. Malanushenko, D. Margala, C. Maraston, K. Mehta, J. Miralda-Escudé, A. D. Myers, R. C. Nichol, P. Noterdaeme, M. D. Olmstead, D. Oravetz, N. Palanque-Delabrouille, K. Pan, I. Pâris, W. J. Percival, P. Petitjean, N. A. Roe, E. Rollinde, N. P. Ross, G. Rossi, D. J. Schlegel, D. P. Schneider, A. Shelden, E. S. Sheldon, A. Simmons, S. Snedden, J. L. Tinker, M. Viel, B. A. Weaver, D. H. Weinberg, M. White, C. Yèche, and D. G. York. Baryon acoustic oscillations in the Lyα forest of BOSS quasars. *A&A*, 552:A96, April 2013. doi: 10.1051/0004-6361/201220724.

M. Cacciato, F. C. van den Bosch, S. More, R. Li, H. J. Mo, and X. Yang. Galaxy clustering and galaxy-galaxy lensing: a promising union to constrain cosmological parameters. *MNRAS*, 394:929–946, April 2009. doi: 10.1111/j.1365-2966.2008.14362.x.

B. Catinella, R. Giovanelli, and M. P. Haynes. Template Rotation Curves for Disk Galaxies. *ApJ*, 640:751–761, April 2006. doi: 10.1086/500171.

B. Catinella, D. Schiminovich, G. Kauffmann, S. Fabello, J. Wang, C. Hummels, J. Lemonias, S. M. Moran, R. Wu, R. Giovanelli, M. P. Haynes, T. M. Heckman,

A. R. Basu-Zych, M. R. Blanton, J. Brinchmann, T. Budavári, T. Gonçalves, B. D. Johnson, R. C. Kennicutt, B. F. Madore, C. D. Martin, M. R. Rich, L. J. Tacconi, D. A. Thilker, V. Wild, and T. K. Wyder. The GALEX Arecibo SDSS Survey - I. Gas fraction scaling relations of massive galaxies and first data release. *MNRAS*, 403:683–708, April 2010. doi: 10.1111/j.1365-2966.2009.16180.x.

A. Cattaneo, A. Dekel, J. Devriendt, B. Guiderdoni, and J. Blaizot. Modelling the galaxy bimodality: shutdown above a critical halo mass. *MNRAS*, 370: 1651–1665, August 2006. doi: 10.1111/j.1365-2966.2006.10608.x.

A. Cattaneo, S. M. Faber, J. Binney, A. Dekel, J. Kormendy, R. Mushotzky, A. Babul, P. N. Best, M. Brüggen, A. C. Fabian, C. S. Frenk, A. Khalatyan, H. Netzer, A. Mahdavi, J. Silk, M. Steinmetz, and L. Wisotzki. The role of black holes in galaxy formation and evolution. *Nature*, 460:213–219, July 2009. doi: 10.1038/nature08135.

G. Chabrier. Galactic Stellar and Substellar Initial Mass Function. *PASP*, 115: 763–795, July 2003. doi: 10.1086/376392.

K.-H. Chae. Galaxy evolution from strong-lensing statistics: the differential evolution of the velocity dispersion function in concord with the Λ cold dark matter paradigm. *MNRAS*, 402:2031–2048, March 2010. doi: 10.1111/j.1365-2966.2009.16073.x.

Y.-Y. Choi, C. Park, and M. S. Vogeley. Internal and Collective Properties of Galaxies in the Sloan Digital Sky Survey. *ApJ*, 658:884–897, April 2007. doi: 10.1086/511060.

L. Christodoulou, C. Eminian, J. Loveday, P. Norberg, I. K. Baldry, P. D. Hurley, S. P. Driver, S. P. Bamford, A. M. Hopkins, J. Liske, J. A. Peacock, J. Bland-Hawthorn, S. Brough, E. Cameron, C. J. Conselice, S. M. Croom, C. S. Frenk, M. Gunawardhana, D. H. Jones, L. S. Kelvin, K. Kuijken, R. C. Nichol, H. Parkinson, K. A. Pimbblet, C. C. Popescu, M. Prescott, A. S. G. Robotham, R. G. Sharp, W. J. Sutherland, E. N. Taylor, D. Thomas, R. J. Tuffs, E. van Kampen, and D. Wijesinghe. Galaxy And Mass Assembly (GAMA): colour- and luminosity-dependent clustering from calibrated photometric redshifts. *MNRAS*, 425:1527–1548, September 2012. doi: 10.1111/j.1365-2966.2012.21434.x.

S. Cole, P. Norberg, C. M. Baugh, C. S. Frenk, J. Bland-Hawthorn, T. Bridges, R. Cannon, M. Colless, C. Collins, W. Couch, N. Cross, G. Dalton, R. De Propris, S. P. Driver, G. Efstathiou, R. S. Ellis, K. Glazebrook, C. Jackson, O. Lahav, I. Lewis, S. Lumsden, S. Maddox, D. Madgwick, J. A. Peacock, B. A. Peterson, W. Sutherland, and K. Taylor. The 2dF galaxy redshift survey: near-infrared galaxy luminosity functions. *MNRAS*, 326:255–273, September 2001. doi: 10.1046/j.1365-8711.2001.04591.x.

M. Colless, G. Dalton, S. Maddox, W. Sutherland, P. Norberg, S. Cole, J. Bland-Hawthorn, T. Bridges, R. Cannon, C. Collins, W. Couch, N. Cross, K. Deeley, R. De Propris, S. P. Driver, G. Efstathiou, R. S. Ellis, C. S. Frenk, K. Glazebrook, C. Jackson, O. Lahav, I. Lewis, S. Lumsden, D. Madgwick, J. A. Peacock, B. A. Peterson, I. Price, M. Seaborne, and K. Taylor. The 2dF Galaxy Redshift Survey: spectra and redshifts. *MNRAS*, 328:1039–1063, December 2001. doi: 10.1046/j.1365-8711.2001.04902.x.

C. Conroy, R. H. Wechsler, and A. V. Kravtsov. Modeling Luminosity-dependent Galaxy Clustering through Cosmic Time. *ApJ*, 647:201–214, August 2006. doi: 10.1086/503602.

C. Conroy, F. Prada, J. A. Newman, D. Croton, A. L. Coil, C. J. Conselice, M. C. Cooper, M. Davis, S. M. Faber, B. F. Gerke, P. Guhathakurta, A. Klypin, D. C. Koo, and R. Yan. Evolution in the Halo Masses of Isolated Galaxies between $z \sim 1$ and $z \sim 0$: From DEEP2 to SDSS. *ApJ*, 654:153–171, January 2007. doi: 10.1086/509632.

C. Contreras, C. Blake, G. B. Poole, F. Marin, S. Brough, M. Colless, W. Couch, S. Croom, D. Croton, T. M. Davis, M. J. Drinkwater, K. Forster, D. Gilbank, M. Gladders, K. Glazebrook, B. Jelliffe, R. J. Jurek, I.-h. Li, B. Madore, D. C. Martin, K. Pimbblet, M. Pracy, R. Sharp, E. Wisnioski, D. Woods, T. K. Wyder, and H. K. C. Yee. The WiggleZ Dark Energy Survey: measuring the cosmic growth rate with the two-point galaxy correlation function. *MNRAS*, 430: 924–933, April 2013. doi: 10.1093/mnras/sts608.

S. Côté, C. Carignan, and K. C. Freeman. The Various Kinematics of Dwarf Irregular Galaxies in Nearby Groups and Their Dark Matter Distributions. *AJ*, 120:3027–3059, December 2000. doi: 10.1086/316883.

D. J. Croton, V. Springel, S. D. M. White, G. De Lucia, C. S. Frenk, L. Gao, A. Jenkins, G. Kauffmann, J. F. Navarro, and N. Yoshida. The many lives of active galactic nuclei: cooling flows, black holes and the luminosities and colours of galaxies. *MNRAS*, 365:11–28, January 2006. doi: 10.1111/j.1365-2966.2005.09675.x.

M. Davis and J. Huchra. A survey of galaxy redshifts. III - The density field and the induced gravity field. *ApJ*, 254:437–450, March 1982. doi: 10.1086/159751.

S. de la Torre, L. Guzzo, J. A. Peacock, E. Branchini, A. Iovino, B. R. Granett, U. Abbas, C. Adami, S. Arnouts, J. Bel, M. Bolzonella, D. Bottini, A. Cappi, J. Coupon, O. Cucciati, I. Davidzon, G. De Lucia, A. Fritz, P. Franzetti, M. Fumana, B. Garilli, O. Ilbert, J. Krywult, V. Le Brun, O. Le Fevre, D. Maccagni, K. Malek, F. Marulli, H. J. McCracken, L. Moscardini, L. Paioro, W. J. Percival, M. Polletta, A. Pollo, H. Schlagenhaufer, M. Scodeggio, L. A. M. Tasca, R. Tojeiro, D. Vergani, A. Zanichelli, A. Burden, C. Di Porto, A. Marchetti, C. Marinoni, Y. Mellier, P. Monaco, R. C. Nichol, S. Phleps, M. Wolk, and G. Zamorani.

The VIMOS Public Extragalactic Redshift Survey (VIPERS). Galaxy clustering and redshift-space distortions at z=0.8 in the first data release. *ArXiv e-prints*, March 2013.

G. De Lucia and J. Blaizot. The hierarchical formation of the brightest cluster galaxies. *MNRAS*, 375:2–14, February 2007. doi: 10.1111/j.1365-2966.2006.11287.x.

A. Dekel and Y. Birnboim. Galaxy bimodality due to cold flows and shock heating. *MNRAS*, 368:2–20, May 2006. doi: 10.1111/j.1365-2966.2006.10145.x.

H. Desmond. The baryonic Tully-Fisher Relation predicted by cold dark matter cosmogony. *ArXiv e-prints*, April 2012.

J. Diemand, M. Kuhlen, and P. Madau. Formation and Evolution of Galaxy Dark Matter Halos and Their Substructure. *ApJ*, 667:859–877, October 2007. doi: 10.1086/520573.

S. Dodelson and L. M. Widrow. Sterile neutrinos as dark matter. *Physical Review Letters*, 72:17–20, January 1994. doi: 10.1103/PhysRevLett.72.17.

M. T. Doyle, M. J. Drinkwater, D. J. Rohde, K. A. Pimbblet, M. Read, M. J. Meyer, M. A. Zwaan, E. Ryan-Weber, J. Stevens, B. S. Koribalski, R. L. Webster, L. Staveley-Smith, D. G. Barnes, M. Howlett, V. A. Kilborn, M. Waugh, M. J. Pierce, R. Bhathal, W. J. G. de Blok, M. J. Disney, R. D. Ekers, K. C. Freeman, D. A. Garcia, B. K. Gibson, J. Harnett, P. A. Henning, H. Jerjen, M. J. Kesteven, P. M. Knezek, S. Mader, M. Marquarding, R. F. Minchin, J. O'Brien, T. Oosterloo, R. M. Price, M. E. Putman, S. D. Ryder, E. M. Sadler, I. M. Stewart, F. Stootman, and A. E. Wright. The HIPASS catalogue - III. Optical counterparts and isolated dark galaxies. *MNRAS*, 361:34–44, July 2005. doi: 10.1111/j.1365-2966.2005.09159.x.

M. J. Drinkwater, R. J. Jurek, C. Blake, D. Woods, K. A. Pimbblet, K. Glazebrook, R. Sharp, M. B. Pracy, S. Brough, M. Colless, W. J. Couch, S. M. Croom, T. M. Davis, D. Forbes, K. Forster, D. G. Gilbank, M. Gladders, B. Jelliffe, N. Jones, I.-H. Li, B. Madore, D. C. Martin, G. B. Poole, T. Small, E. Wisnioski, T. Wyder, and H. K. C. Yee. The WiggleZ Dark Energy Survey: survey design and first data release. *MNRAS*, 401:1429–1452, January 2010. doi: 10.1111/j.1365-2966.2009.15754.x.

S. P. Driver and A. S. G. Robotham. Quantifying cosmic variance. *MNRAS*, 407: 2131–2140, October 2010. doi: 10.1111/j.1365-2966.2010.17028.x.

A. A. Dutton. The baryonic Tully-Fisher relation and galactic outflows. *MNRAS*, 424:3123–3128, August 2012. doi: 10.1111/j.1365-2966.2012.21469.x.

A. A. Dutton, C. Conroy, F. C. van den Bosch, F. Prada, and S. More. The kinematic connection between galaxies and dark matter haloes. *MNRAS*, 407: 2–16, September 2010. doi: 10.1111/j.1365-2966.2010.16911.x.

G. Efstathiou, R. S. Ellis, and B. A. Peterson. Analysis of a complete galaxy redshift survey. II - The field-galaxy luminosity function. *MNRAS*, 232:431–461, May 1988.

D. J. Eisenstein, I. Zehavi, D. W. Hogg, R. Scoccimarro, M. R. Blanton, R. C. Nichol, R. Scranton, H.-J. Seo, M. Tegmark, Z. Zheng, S. F. Anderson, J. Annis, N. Bahcall, J. Brinkmann, S. Burles, F. J. Castander, A. Connolly, I. Csabai, M. Doi, M. Fukugita, J. A. Frieman, K. Glazebrook, J. E. Gunn, J. S. Hendry, G. Hennessy, Z. Ivezić, S. Kent, G. R. Knapp, H. Lin, Y.-S. Loh, R. H. Lupton, B. Margon, T. A. McKay, A. Meiksin, J. A. Munn, A. Pope, M. W. Richmond, D. Schlegel, D. P. Schneider, K. Shimasaku, C. Stoughton, M. A. Strauss, M. SubbaRao, A. S. Szalay, I. Szapudi, D. L. Tucker, B. Yanny, and D. G. York. Detection of the Baryon Acoustic Peak in the Large-Scale Correlation Function of SDSS Luminous Red Galaxies. *ApJ*, 633:560–574, November 2005. doi: 10.1086/466512.

C. Evoli, P. Salucci, A. Lapi, and L. Danese. The H I Content of Local Late-type Galaxies. *ApJ*, 743:45, December 2011. doi: 10.1088/0004-637X/743/1/45.

I. Ferrero, M. G. Abadi, J. F. Navarro, L. V. Sales, and S. Gurovich. The dark matter haloes of dwarf galaxies: a challenge for the Λ cold dark matter paradigm? *MNRAS*, 425:2817–2823, October 2012. doi: 10.1111/j.1365-2966.2012.21623.x.

A. S. Font, R. G. Bower, I. G. McCarthy, A. J. Benson, C. S. Frenk, J. C. Helly, C. G. Lacey, C. M. Baugh, and S. Cole. The colours of satellite galaxies in groups and clusters. *MNRAS*, 389:1619–1629, October 2008. doi: 10.1111/j.1365-2966.2008.13698.x.

G. M. Fuller, A. Kusenko, I. Mocioiu, and S. Pascoli. Pulsar kicks from a dark-matter sterile neutrino. *Phys. Rev. D*, 68(10):103002–+, November 2003. doi: 10.1103/PhysRevD.68.103002.

A. Gallazzi, S. Charlot, J. Brinchmann, S. D. M. White, and C. A. Tremonti. The ages and metallicities of galaxies in the local universe. *MNRAS*, 362:41–58, September 2005. doi: 10.1111/j.1365-2966.2005.09321.x.

D. R. Garnett. The Luminosity-Metallicity Relation, Effective Yields, and Metal Loss in Spiral and Irregular Galaxies. *ApJ*, 581:1019–1031, December 2002. doi: 10.1086/344301.

M. Geha, M. R. Blanton, M. Masjedi, and A. A. West. The Baryon Content of Extremely Low Mass Dwarf Galaxies. *ApJ*, 653:240–254, December 2006. doi: 10.1086/508604.

R. Giovanelli and M. P. Haynes. Gas deficiency in cluster galaxies - A comparison of nine clusters. *ApJ*, 292:404–425, May 1985. doi: 10.1086/163170.

R. Giovanelli, M. P. Haynes, B. R. Kent, P. Perillat, B. Catinella, G. L. Hoffman, E. Momjian, J. L. Rosenberg, A. Saintonge, K. Spekkens, S. Stierwalt, N. Brosch, K. L. Masters, C. M. Springob, I. D. Karachentsev, V. E. Karachentseva, R. A. Koopmann, E. Muller, W. van Driel, and L. van Zee. The Arecibo Legacy Fast ALFA Survey. II. Results of Precursor Observations. *AJ*, 130:2613–2624, December 2005. doi: 10.1086/497432.

R. Giovanelli, M. P. Haynes, B. R. Kent, A. Saintonge, S. Stierwalt, A. Altaf, T. Balonek, N. Brosch, S. Brown, B. Catinella, A. Furniss, J. Goldstein, G. L. Hoffman, R. A. Koopmann, D. A. Kornreich, B. Mahmood, A. M. Martin, K. L. Masters, A. Mitschang, E. Momjian, P. H. Nair, J. L. Rosenberg, and B. Walsh. The Arecibo Legacy Fast ALFA Survey. III. H I Source Catalog of the Northern Virgo Cluster Region. *AJ*, 133:2569–2583, June 2007. doi: 10.1086/516635.

K. Glazebrook, J. Tober, S. Thomson, J. Bland-Hawthorn, and R. Abraham. Cosmic Star Formation History to z=1 from a Narrow Emission Line-selected Tunable-Filter Survey. *AJ*, 128:2652–2659, December 2004. doi: 10.1086/425552.

A. H. Gonzalez, K. A. Williams, J. S. Bullock, T. S. Kolatt, and J. R. Primack. The Velocity Function of Galaxies. *ApJ*, 528:145–155, January 2000. doi: 10.1086/308159.

J. Guedes, S. Callegari, P. Madau, and L. Mayer. Forming Realistic Late-type Spirals in a ΛCDM Universe: The Eris Simulation. *ApJ*, 742:76, December 2011. doi: 10.1088/0004-637X/742/2/76.

Q. Guo, S. White, C. Li, and M. Boylan-Kolchin. How do galaxies populate dark matter haloes? *MNRAS*, 404:1111–1120, May 2010. doi: 10.1111/j.1365-2966.2010.16341.x.

Q. Guo, S. White, M. Boylan-Kolchin, G. De Lucia, G. Kauffmann, G. Lemson, C. Li, V. Springel, and S. Weinmann. From dwarf spheroidals to cD galaxies: simulating the galaxy population in a ΛCDM cosmology. *MNRAS*, 413:101–131, May 2011. doi: 10.1111/j.1365-2966.2010.18114.x.

M. Hall, S. Courteau, A. A. Dutton, M. McDonald, and Y. Zhu. An investigation of Sloan Digital Sky Survey imaging data and multiband scaling relations of spiral galaxies. *MNRAS*, 425:2741–2765, October 2012. doi: 10.1111/j.1365-2966.2012.21290.x.

G. Hallenbeck, E. Papastergis, S. Huang, M. P. Haynes, R. Giovanelli, A. Boselli, S. Boissier, S. Heinis, L. Cortese, and S. Fabello. Gas-bearing Early-type Dwarf

Galaxies in Virgo: Evidence for Recent Accretion. *AJ*, 144:87, September 2012. doi: 10.1088/0004-6256/144/3/87.

M. P. Haynes and R. Giovanelli. The pattern of H I deficiency in the Virgo cluster. *ApJ*, 306:466–482, July 1986. doi: 10.1086/164357.

M. P. Haynes, R. Giovanelli, A. M. Martin, K. M. Hess, A. Saintonge, E. A. K. Adams, G. Hallenbeck, G. L. Hoffman, S. Huang, B. R. Kent, R. A. Koopmann, E. Papastergis, S. Stierwalt, T. J. Balonek, D. W. Craig, S. J. U. Higdon, D. A. Kornreich, J. R. Miller, A. A. O'Donoghue, R. P. Olowin, J. L. Rosenberg, K. Spekkens, P. Troischt, and E. M. Wilcots. The Arecibo Legacy Fast ALFA Survey: The α.40 H I Source Catalog, Its Characteristics and Their Impact on the Derivation of the H I Mass Function. *AJ*, 142:170, November 2011. doi: 10.1088/0004-6256/142/5/170.

A. P. Hearin and D. F. Watson. The Dark Side of Galaxy Color. *ArXiv e-prints*, April 2013.

A. P. Hearin, A. R. Zentner, A. A. Berlind, and J. A. Newman. SHAM Beyond Clustering: New Tests of Galaxy-Halo Abundance Matching with Galaxy Groups. *ArXiv e-prints*, October 2012.

M. Hoeft, G. Yepes, S. Gottlöber, and V. Springel. Dwarf galaxies in voids: suppressing star formation with photoheating. *MNRAS*, 371:401–414, September 2006. doi: 10.1111/j.1365-2966.2006.10678.x.

H. Hoekstra. A comparison of weak-lensing masses and X-ray properties of galaxy clusters. *MNRAS*, 379:317–330, July 2007. doi: 10.1111/j.1365-2966.2007.11951.x.

S. Huang, M. P. Haynes, R. Giovanelli, and J. Brinchmann. The Arecibo Legacy Fast ALFA Survey: The Galaxy Population Detected by ALFALFA. *ApJ*, 756: 113, September 2012a. doi: 10.1088/0004-637X/756/2/113.

S. Huang, M. P. Haynes, R. Giovanelli, J. Brinchmann, S. Stierwalt, and S. G. Neff. Gas, Stars, and Star Formation in ALFALFA Dwarf Galaxies. *AJ*, 143: 133, June 2012b. doi: 10.1088/0004-6256/143/6/133.

N. Jarosik, C. L. Bennett, J. Dunkley, B. Gold, M. R. Greason, M. Halpern, R. S. Hill, G. Hinshaw, A. Kogut, E. Komatsu, D. Larson, M. Limon, S. S. Meyer, M. R. Nolta, N. Odegard, L. Page, K. M. Smith, D. N. Spergel, G. S. Tucker, J. L. Weiland, E. Wollack, and E. L. Wright. Seven-year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Sky Maps, Systematic Errors, and Basic Results. *ApJS*, 192:14, February 2011. doi: 10.1088/0067-0049/192/2/14.

I. D. Karachentsev, V. E. Karachentseva, W. K. Huchtmeier, and D. I. Makarov. A Catalog of Neighboring Galaxies. *AJ*, 127:2031–2068, April 2004. doi: 10.1086/382905.

G. Kauffmann, T. M. Heckman, S. D. M. White, S. Charlot, C. Tremonti, E. W. Peng, M. Seibert, J. Brinkmann, R. C. Nichol, M. SubbaRao, and D. York. The dependence of star formation history and internal structure on stellar mass for 10^5 low-redshift galaxies. *MNRAS*, 341:54–69, May 2003. doi: 10.1046/j.1365-8711.2003.06292.x.

B. R. Kent, R. Giovanelli, M. P. Haynes, A. M. Martin, A. Saintonge, S. Stierwalt, T. J. Balonek, N. Brosch, and R. A. Koopmann. The Arecibo Legacy Fast Alfa Survey. VI. Second HI Source Catalog of the Virgo Cluster Region. *AJ*, 136: 713–724, August 2008. doi: 10.1088/0004-6256/136/2/713.

D. Kereš, N. Katz, D. H. Weinberg, and R. Davé. How do galaxies get their gas? *MNRAS*, 363:2–28, October 2005. doi: 10.1111/j.1365-2966.2005.09451.x.

J.-h. Kim and J. Lee. How Does the Surface Density and Size of Disk Galaxies Measured in Hydrodynamic Simulations Correlate with the Halo Spin Parameter? *ArXiv e-prints*, October 2012.

A. Klypin and J. Holtzman. Particle-Mesh code for cosmological simulations. *ArXiv Astrophysics e-prints*, December 1997.

A. Klypin, A. V. Kravtsov, O. Valenzuela, and F. Prada. Where Are the Missing Galactic Satellites? *ApJ*, 522:82–92, September 1999. doi: 10.1086/307643.

A. Klypin, H. Zhao, and R. S. Somerville. ΛCDM-based Models for the Milky Way and M31. I. Dynamical Models. *ApJ*, 573:597–613, July 2002. doi: 10.1086/340656.

A. Klypin, S. Trujillo-Gomez, and J. Primack. Halos and galaxies in the standard cosmological model: results from the Bolshoi simulation. *ArXiv e-prints*, February 2010.

A. A. Klypin, S. Trujillo-Gomez, and J. Primack. Dark Matter Halos in the Standard Cosmological Model: Results from the Bolshoi Simulation. *ApJ*, 740:102, October 2011. doi: 10.1088/0004-637X/740/2/102.

E. Komatsu, K. M. Smith, J. Dunkley, C. L. Bennett, B. Gold, G. Hinshaw, N. Jarosik, D. Larson, M. R. Nolta, L. Page, D. N. Spergel, M. Halpern, R. S. Hill, A. Kogut, M. Limon, S. S. Meyer, N. Odegard, G. S. Tucker, J. L. Weiland, E. Wollack, and E. L. Wright. Seven-year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Cosmological Interpretation. *ApJS*, 192:18–+, February 2011. doi: 10.1088/0067-0049/192/2/18.

B. S. Koribalski. Overview on Spectral Line Source Finding and Visualisation. *PASA*, 29:359–370, August 2012. doi: 10.1071/AS12030.

A. Kusenko. Sterile neutrinos: The dark side of the light fermions. *Phys. Rep.*, 481:1–28, September 2009. doi: 10.1016/j.physrep.2009.07.004.

S. D. Landy and A. S. Szalay. Bias and variance of angular correlation functions. *ApJ*, 412:64–71, July 1993. doi: 10.1086/172900.

A. Leauthaud, A. Finoguenov, J.-P. Kneib, J. E. Taylor, R. Massey, J. Rhodes, O. Ilbert, K. Bundy, J. Tinker, M. R. George, P. Capak, A. M. Koekemoer, D. E. Johnston, Y.-Y. Zhang, N. Cappelluti, R. S. Ellis, M. Elvis, S. Giodini, C. Heymans, O. Le Fèvre, S. Lilly, H. J. McCracken, Y. Mellier, A. Réfrégier, M. Salvato, N. Scoville, G. Smoot, M. Tanaka, L. Van Waerbeke, and M. Wolk. A Weak Lensing Study of X-ray Groups in the Cosmos Survey: Form and Evolution of the Mass-Luminosity Relation. *ApJ*, 709:97–114, January 2010. doi: 10.1088/0004-637X/709/1/97.

A. Leauthaud, J. Tinker, K. Bundy, P. S. Behroozi, R. Massey, J. Rhodes, M. R. George, J.-P. Kneib, A. Benson, R. H. Wechsler, M. T. Busha, P. Capak, M. Cortês, O. Ilbert, A. M. Koekemoer, O. Le Fèvre, S. Lilly, H. J. McCracken, M. Salvato, T. Schrabback, N. Scoville, T. Smith, and J. E. Taylor. New Constraints on the Evolution of the Stellar-to-dark Matter Connection: A Combined Analysis of Galaxy-Galaxy Lensing, Clustering, and Stellar Mass Functions from z = 0.2 to z =1. *ApJ*, 744:159, January 2012. doi: 10.1088/0004-637X/744/2/159.

A. K. Leroy, F. Walter, F. Bigiel, A. Usero, A. Weiss, E. Brinks, W. J. G. de Blok, R. C. Kennicutt, K.-F. Schuster, C. Kramer, H. W. Wiesemeyer, and H. Roussel. Heracles: The HERA CO Line Extragalactic Survey. *AJ*, 137:4670–4696, June 2009. doi: 10.1088/0004-6256/137/6/4670.

C. Li and S. D. M. White. The distribution of stellar mass in the low-redshift Universe. *MNRAS*, 398:2177–2187, October 2009. doi: 10.1111/j.1365-2966.2009.15268.x.

C. Li, G. Kauffmann, Y. P. Jing, S. D. M. White, G. Börner, and F. Z. Cheng. The dependence of clustering on galaxy properties. *MNRAS*, 368:21–36, May 2006a. doi: 10.1111/j.1365-2966.2006.10066.x.

C. Li, G. Kauffmann, L. Wang, S. D. M. White, T. M. Heckman, and Y. P. Jing. The clustering of narrow-line AGN in the local Universe. *MNRAS*, 373:457–468, December 2006b. doi: 10.1111/j.1365-2966.2006.11079.x.

C. Li, G. Kauffmann, J. Fu, J. Wang, B. Catinella, S. Fabello, D. Schiminovich, and W. Zhang. The clustering of galaxies as a function of their photometrically estimated atomic gas content. *MNRAS*, 424:1471–1482, August 2012a. doi: 10.1111/j.1365-2966.2012.21337.x.

C. Li, S. D. M. White, Y. Chen, A. L. Coil, M. Davis, G. De Lucia, Q. Guo, Y. P. Jing, G. Kauffmann, C. N. A. Willmer, and W. Zhang. Autocorrelations of stellar light and mass at $z \sim 0$ and ~ 1: from SDSS to DEEP2. *MNRAS*, 419: 1557–1565, January 2012b. doi: 10.1111/j.1365-2966.2011.19817.x.

Y. Lu, H. J. Mo, N. Katz, and M. D. Weinberg. Bayesian inference of galaxy formation from the K-band luminosity function of galaxies: tensions between theory and observation. *MNRAS*, 421:1779–1796, April 2012. doi: 10.1111/j.1365-2966.2012.20435.x.

A. V. Macciò, A. A. Dutton, and F. C. van den Bosch. Concentration, spin and shape of dark matter haloes as a function of the cosmological model: WMAP1, WMAP3 and WMAP5 results. *MNRAS*, 391:1940–1954, December 2008. doi: 10.1111/j.1365-2966.2008.14029.x.

R. Mandelbaum, U. Seljak, G. Kauffmann, C. M. Hirata, and J. Brinkmann. Galaxy halo masses and satellite fractions from galaxy-galaxy lensing in the Sloan Digital Sky Survey: stellar mass, luminosity, morphology and environment dependencies. *MNRAS*, 368:715–731, May 2006. doi: 10.1111/j.1365-2966.2006.10156.x.

F. A. Marín, N. Y. Gnedin, H.-J. Seo, and A. Vallinotto. Modeling the Large-scale Bias of Neutral Hydrogen. *ApJ*, 718:972–980, August 2010. doi: 10.1088/0004-637X/718/2/972.

C. Marinoni and M. J. Hudson. The Mass-to-Light Function of Virialized Systems and the Relationship between Their Optical and X-Ray Properties. *ApJ*, 569:101–111, April 2002. doi: 10.1086/339319.

A. M. Martin, R. Giovanelli, M. P. Haynes, A. Saintonge, G. L. Hoffman, B. R. Kent, and S. Stierwalt. The Arecibo Legacy Fast Alfa Survey. VIII. H I Source Catalog of the Anti-Virgo Region at $\delta = +25^\circ$. *ApJS*, 183:214–224, August 2009. doi: 10.1088/0067-0049/183/2/214.

A. M. Martin, E. Papastergis, R. Giovanelli, M. P. Haynes, C. M. Springob, and S. Stierwalt. The Arecibo Legacy Fast ALFA Survey. X. The H I Mass Function and \Omega _H\,{\mathsc{i}} from the 40% ALFALFA Survey. *ApJ*, 723: 1359–1374, November 2010. doi: 10.1088/0004-637X/723/2/1359.

A. M. Martin, R. Giovanelli, M. P. Haynes, and L. Guzzo. The Clustering Characteristics of H I-selected Galaxies from the 40% ALFALFA Survey. *ApJ*, 750: 38, May 2012. doi: 10.1088/0004-637X/750/1/38.

F. Marulli, M. Bolzonella, E. Branchini, I. Davidzon, S. de la Torre, B. R. Granett, L. Guzzo, A. Iovino, L. Moscardini, A. Pollo, U. Abbas, C. Adami, S. Arnouts, J. Bel, D. Bottini, A. Cappi, J. Coupon, O. Cucciati, G. De Lucia, A. Fritz, P. Franzetti, M. Fumana, B. Garilli, O. Ilbert, J. Krywult, V. Le Brun, O. Le Fevre, D. Maccagni, K. Malek, H. J. McCracken, L. Paioro, M. Polletta, H. Schlagenhaufer, M. Scodeggio, L. A. M. Tasca, R. Tojeiro, D. Vergani, A. Zanichelli, A. Burden, C. Di Porto, A. Marchetti, C. Marinoni, Y. Mellier, R. C. Nichol, J. A. Peacock, W. J. Percival, S. Phleps, M. Wolk, and

G. Zamorani. The VIMOS Public Extragalactic Redshift Survey (VIPERS). Luminosity and stellar mass dependence of galaxy clustering at $0.5 < z < 1.1$. *ArXiv e-prints*, March 2013.

K. L. Masters. *Galaxy flows in and around the Local Supercluster*. PhD thesis, Cornell University, United States – New York, 2005.

K. L. Masters, M. P. Haynes, and R. Giovanelli. The Impact of Distance Uncertainties on Local Luminosity and Mass Functions. *ApJL*, 607:L115–L118, June 2004. doi: 10.1086/422100.

K. W. Masui, E. R. Switzer, N. Banavar, K. Bandura, C. Blake, L.-M. Calin, T.-C. Chang, X. Chen, Y.-C. Li, Y.-W. Liao, A. Natarajan, U.-L. Pen, J. B. Peterson, J. R. Shaw, and T. C. Voytek. Measurement of 21 cm Brightness Fluctuations at $z \sim 0.8$ in Cross-correlation. *ApJL*, 763:L20, January 2013. doi: 10.1088/2041-8205/763/1/L20.

I. G. McCarthy, J. Schaye, A. S. Font, T. Theuns, C. S. Frenk, R. A. Crain, and C. Dalla Vecchia. Rotation rates, sizes and star formation efficiencies of a representative population of simulated disc galaxies. *MNRAS*, 427:379–392, November 2012. doi: 10.1111/j.1365-2966.2012.21951.x.

S. S. McGaugh. The Baryonic Tully-Fisher Relation of Galaxies with Extended Rotation Curves and the Stellar Mass of Rotating Galaxies. *ApJ*, 632:859–871, October 2005. doi: 10.1086/432968.

S. S. McGaugh. The Baryonic Tully-Fisher Relation of Gas-rich Galaxies as a Test of ΛCDM and MOND. *AJ*, 143:40, February 2012. doi: 10.1088/0004-6256/143/2/40.

M. J. Meyer, M. A. Zwaan, R. L. Webster, L. Staveley-Smith, E. Ryan-Weber, M. J. Drinkwater, D. G. Barnes, M. Howlett, V. A. Kilborn, J. Stevens, M. Waugh, M. J. Pierce, R. Bhathal, W. J. G. de Blok, M. J. Disney, R. D. Ekers, K. C. Freeman, D. A. Garcia, B. K. Gibson, J. Harnett, P. A. Henning, H. Jerjen, M. J. Kesteven, P. M. Knezek, B. S. Koribalski, S. Mader, M. Marquarding, R. F. Minchin, J. O'Brien, T. Oosterloo, R. M. Price, M. E. Putman, S. D. Ryder, E. M. Sadler, I. M. Stewart, F. Stootman, and A. E. Wright. The HIPASS catalogue - I. Data presentation. *MNRAS*, 350:1195–1209, June 2004. doi: 10.1111/j.1365-2966.2004.07710.x.

M. J. Meyer, M. A. Zwaan, R. L. Webster, M. J. I. Brown, and L. Staveley-Smith. The Weak Clustering of Gas-rich Galaxies. *ApJ*, 654:702–713, January 2007. doi: 10.1086/508799.

H. J. Mo, S. Mao, and S. D. M. White. The formation of galactic discs. *MNRAS*, 295:319–336, April 1998. doi: 10.1046/j.1365-8711.1998.01227.x.

A. D. Montero-Dorta and F. Prada. The SDSS DR6 luminosity functions of galaxies. *MNRAS*, 399:1106–1118, November 2009. doi: 10.1111/j.1365-2966.2009.15197.x.

B. Moore, S. Ghigna, F. Governato, G. Lake, T. Quinn, J. Stadel, and P. Tozzi. Dark Matter Substructure within Galactic Halos. *ApJL*, 524:L19–L22, October 1999. doi: 10.1086/312287.

S. More, F. C. van den Bosch, M. Cacciato, H. J. Mo, X. Yang, and R. Li. Satellite kinematics - II. The halo mass-luminosity relation of central galaxies in SDSS. *MNRAS*, 392:801–816, January 2009. doi: 10.1111/j.1365-2966.2008.14095.x.

S. More, F. C. van den Bosch, M. Cacciato, R. Skibba, H. J. Mo, and X. Yang. Satellite kinematics - III. Halo masses of central galaxies in SDSS. *MNRAS*, 410:210–226, January 2011. doi: 10.1111/j.1365-2966.2010.17436.x.

B. P. Moster, R. S. Somerville, C. Maulbetsch, F. C. van den Bosch, A. V. Macciò, T. Naab, and L. Oser. Constraints on the Relationship between Stellar Mass and Halo Mass at Low and High Redshift. *ApJ*, 710:903–923, February 2010. doi: 10.1088/0004-637X/710/2/903.

M. Musso, A. Paranjape, and R. K. Sheth. Scale-dependent halo bias in the excursion set approach. *MNRAS*, 427:3145–3158, December 2012. doi: 10.1111/j.1365-2966.2012.21903.x.

S. F. Newman, K. Shapiro Griffin, R. Genzel, R. Davies, N. M. Förster-Schreiber, L. J. Tacconi, J. Kurk, S. Wuyts, S. Genel, S. J. Lilly, A. Renzini, N. Bouché, A. Burkert, G. Cresci, P. Buschkamp, C. M. Carollo, F. Eisenhauer, E. Hicks, D. Lutz, C. Mancini, T. Naab, Y. Peng, and D. Vergani. Shocked Superwinds from the z ˜ 2 Clumpy Star-forming Galaxy, ZC406690. *ApJ*, 752:111, June 2012. doi: 10.1088/0004-637X/752/2/111.

E. Noordermeer, J. M. van der Hulst, R. Sancisi, R. A. Swaters, and T. S. van Albada. The Westerbork HI survey of spiral and irregular galaxies. III. HI observations of early-type disk galaxies. *A&A*, 442:137–157, October 2005. doi: 10.1051/0004-6361:20053172.

P. Norberg, C. M. Baugh, E. Hawkins, S. Maddox, D. Madgwick, O. Lahav, S. Cole, C. S. Frenk, I. Baldry, J. Bland-Hawthorn, T. Bridges, R. Cannon, M. Colless, C. Collins, W. Couch, G. Dalton, R. De Propris, S. P. Driver, G. Efstathiou, R. S. Ellis, K. Glazebrook, C. Jackson, I. Lewis, S. Lumsden, J. A. Peacock, B. A. Peterson, W. Sutherland, and K. Taylor. The 2dF Galaxy Redshift Survey: the dependence of galaxy clustering on luminosity and spectral type. *MNRAS*, 332:827–838, June 2002. doi: 10.1046/j.1365-8711.2002.05348.x.

D. Obreschkow, D. Croton, G. De Lucia, S. Khochfar, and S. Rawlings. Simulation of the Cosmic Evolution of Atomic and Molecular Hydrogen in Galaxies. *ApJ*, 698:1467–1484, June 2009. doi: 10.1088/0004-637X/698/2/1467.

D. Obreschkow, X. Ma, M. Meyer, C. Power, M. Zwaan, L. Staveley-Smith, and M. J. Drinkwater. Confronting Cold Dark Matter Predictions with Observed Galaxy Rotations. *ApJ*, 766:137, April 2013. doi: 10.1088/0004-637X/766/2/137.

S.-H. Oh, W. J. G. de Blok, E. Brinks, F. Walter, and R. C. Kennicutt, Jr. Dark and Luminous Matter in THINGS Dwarf Galaxies. *AJ*, 141:193, June 2011. doi: 10.1088/0004-6256/141/6/193.

T. Okamoto, L. Gao, and T. Theuns. Mass loss of galaxies due to an ultraviolet background. *MNRAS*, 390:920–928, November 2008. doi: 10.1111/j.1365-2966.2008.13830.x.

B. Panter, R. Jimenez, A. F. Heavens, and S. Charlot. The star formation histories of galaxies in the Sloan Digital Sky Survey. *MNRAS*, 378:1550–1564, July 2007. doi: 10.1111/j.1365-2966.2007.11909.x.

E. Papastergis, A. M. Martin, R. Giovanelli, and M. P. Haynes. The Velocity Width Function of Galaxies from the 40% ALFALFA Survey: Shedding Light on the Cold Dark Matter Overabundance Problem. *ApJ*, 739:38, September 2011. doi: 10.1088/0004-637X/739/1/38.

E. Papastergis, A. Cattaneo, S. Huang, R. Giovanelli, and M. P. Haynes. A Direct Measurement of the Baryonic Mass Function of Galaxies and Implications for the Galactic Baryon Fraction. *ApJ*, 759:138, November 2012. doi: 10.1088/0004-637X/759/2/138.

J. Peñarrubia, J. F. Navarro, and A. W. McConnachie. The Tidal Evolution of Local Group Dwarf Spheroidals. *ApJ*, 673:226–240, January 2008. doi: 10.1086/523686.

P. J. E. Peebles. The Void Phenomenon. *ApJ*, 557:495–504, August 2001. doi: 10.1086/322254.

J. Pforr, C. Maraston, and C. Tonini. Recovering galaxy stellar population properties from broad-band spectral energy distribution fitting. *MNRAS*, 422: 3285–3326, June 2012. doi: 10.1111/j.1365-2966.2012.20848.x.

J. Pizagno, F. Prada, D. H. Weinberg, H.-W. Rix, R. W. Pogge, E. K. Grebel, D. Harbeck, M. Blanton, J. Brinkmann, and J. E. Gunn. The Tully-Fisher Relation and its Residuals for a Broadly Selected Sample of Galaxies. *AJ*, 134: 945–972, September 2007. doi: 10.1086/519522.

W. H. Press and P. Schechter. Formation of Galaxies and Clusters of Galaxies by Self-Similar Gravitational Condensation. *ApJ*, 187:425–438, February 1974. doi: 10.1086/152650.

R. M. Reddick, R. H. Wechsler, J. L. Tinker, and P. S. Behroozi. The Connection between Galaxies and Dark Matter Structures in the Local Universe. *ArXiv e-prints*, July 2012.

B. A. Reid, L. Samushia, M. White, W. J. Percival, M. Manera, N. Padmanabhan, A. J. Ross, A. G. Sánchez, S. Bailey, D. Bizyaev, A. S. Bolton, H. Brewington, J. Brinkmann, J. R. Brownstein, A. J. Cuesta, D. J. Eisenstein, J. E. Gunn, K. Honscheid, E. Malanushenko, V. Malanushenko, C. Maraston, C. K. McBride, D. Muna, R. C. Nichol, D. Oravetz, K. Pan, R. de Putter, N. A. Roe, N. P. Ross, D. J. Schlegel, D. P. Schneider, H.-J. Seo, A. Shelden, E. S. Sheldon, A. Simmons, R. A. Skibba, S. Snedden, M. E. C. Swanson, D. Thomas, J. Tinker, R. Tojeiro, L. Verde, D. A. Wake, B. A. Weaver, D. H. Weinberg, I. Zehavi, and G.-B. Zhao. The clustering of galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey: measurements of the growth of structure and expansion rate at z = 0.57 from anisotropic clustering. *MNRAS*, 426:2719–2737, November 2012. doi: 10.1111/j.1365-2966.2012.21779.x.

R. Reyes, R. Mandelbaum, J. E. Gunn, R. Nakajima, U. Seljak, and C. M. Hirata. Optical-to-virial velocity ratios of local disc galaxies from combined kinematics and galaxy-galaxy lensing. *MNRAS*, 425:2610–2640, October 2012. doi: 10.1111/j.1365-2966.2012.21472.x.

R. J. Reynolds. Warm ionized gas in the local interstellar medium. *Advances in Space Research*, 34:27–34, January 2004. doi: 10.1016/j.asr.2003.02.059.

M. Ricotti. The First Galaxies and the Likely Discovery of Their Fossils in the Local Group. *Advances in Astronomy*, 2010, 2010. doi: 10.1155/2010/271592.

A. Rodríguez-Puebla, V. Avila-Reese, C. Firmani, and P. Colín. On the stellar and baryonic mass fractions of central blue and red galaxies. *Rev.Mex.A.A.*, 47:235–253, October 2011.

A. Rodríguez-Puebla, N. Drory, and V. Avila-Reese. The Stellar-Subhalo Mass Relation of Satellite Galaxies. *ApJ*, 756:2, September 2012. doi: 10.1088/0004-637X/756/1/2.

A. Rodríguez-Puebla, V. Avila-Reese, and N. Drory. The Galaxy-Halo/Subhalo Connection: Mass Relations and Implications for Some Satellite Occupational Distributions. *ApJ*, 767:92, April 2013. doi: 10.1088/0004-637X/767/1/92.

J. L. Rosenberg and S. E. Schneider. The Arecibo Dual-Beam Survey: The H I Mass Function of Galaxies. *ApJ*, 567:247–257, March 2002. doi: 10.1086/338377.

A. Saintonge. The Arecibo Legacy Fast ALFA Survey. IV. Strategies for Signal Identification and Survey Catalog Reliability. *AJ*, 133:2087–2096, May 2007. doi: 10.1086/513515.

A. Saintonge, R. Giovanelli, M. P. Haynes, G. L. Hoffman, B. R. Kent, A. M. Martin, S. Stierwalt, and N. Brosch. The Arecibo Legacy Fast Alfa Survey. V. The H I Source Catalog of the Anti-Virgo Region at $\delta = +27^\circ$. *AJ*, 135:588–604, February 2008. doi: 10.1088/0004-6256/135/2/588.

A. Saintonge, G. Kauffmann, C. Kramer, L. J. Tacconi, C. Buchbender, B. Catinella, S. Fabello, J. Graciá-Carpio, J. Wang, L. Cortese, J. Fu, R. Genzel, R. Giovanelli, Q. Guo, M. P. Haynes, T. M. Heckman, M. R. Krumholz, J. Lemonias, C. Li, S. Moran, N. Rodriguez-Fernandez, D. Schiminovich, K. Schuster, and A. Sievers. COLD GASS, an IRAM legacy survey of molecular gas in massive galaxies - I. Relations between H_2, H I, stellar content and structural properties. *MNRAS*, 415:32–60, July 2011. doi: 10.1111/j.1365-2966.2011.18677.x.

S. Salim, R. M. Rich, S. Charlot, J. Brinchmann, B. D. Johnson, D. Schiminovich, M. Seibert, R. Mallery, T. M. Heckman, K. Forster, P. G. Friedman, D. C. Martin, P. Morrissey, S. G. Neff, T. Small, T. K. Wyder, L. Bianchi, J. Donas, Y.-W. Lee, B. F. Madore, B. Milliard, A. S. Szalay, B. Y. Welsh, and S. K. Yi. UV Star Formation Rates in the Local Universe. *ApJS*, 173:267–292, December 2007. doi: 10.1086/519218.

P. Salucci and M. Persic. The baryonic mass function of spiral galaxies: clues to galaxy formation. *MNRAS*, 309:923–928, November 1999. doi: 10.1046/j.1365-8711.1999.02913.x.

A. Sandage, G. A. Tammann, and A. Yahil. The velocity field of bright nearby galaxies. I - The variation of mean absolute magnitude with redshift for galaxies in a magnitude-limited sample. *ApJ*, 232:352–364, September 1979. doi: 10.1086/157295.

M. Schmidt. Space Distribution and Luminosity Functions of Quasi-Stellar Radio Sources. *ApJ*, 151:393–+, February 1968. doi: 10.1086/149446.

F. Shankar, A. Lapi, P. Salucci, G. De Zotti, and L. Danese. New Relationships between Galaxy Properties and Host Halo Mass, and the Role of Feedbacks in Galaxy Formation. *ApJ*, 643:14–25, May 2006. doi: 10.1086/502794.

R. K. Sheth and G. Tormen. An excursion set model of hierarchical clustering: ellipsoidal collapse and the moving barrier. *MNRAS*, 329:61–75, January 2002. doi: 10.1046/j.1365-8711.2002.04950.x.

J. D. Simon and M. Geha. The Kinematics of the Ultra-faint Milky Way Satellites: Solving the Missing Satellite Problem. *ApJ*, 670:313–331, November 2007. doi: 10.1086/521816.

M. C. Smith, G. R. Ruchti, A. Helmi, R. F. G. Wyse, J. P. Fulbright, K. C. Freeman, J. F. Navarro, G. M. Seabroke, M. Steinmetz, M. Williams, O. Bienaymé, J. Binney, J. Bland-Hawthorn, W. Dehnen, B. K. Gibson, G. Gilmore, E. K. Grebel, U. Munari, Q. A. Parker, R.-D. Scholz, A. Siebert, F. G. Watson, and T. Zwitter. The RAVE survey: constraining the local Galactic escape speed. *MNRAS*, 379: 755–772, August 2007. doi: 10.1111/j.1365-2966.2007.11964.x.

J. M. Solanes, T. Sanchis, E. Salvador-Solé, R. Giovanelli, and M. P. Haynes. The Three-dimensional Structure of the Virgo Cluster Region from Tully-Fisher and H I Data. *AJ*, 124:2440–2452, November 2002. doi: 10.1086/344074.

R. S. Somerville, K. Lee, H. C. Ferguson, J. P. Gardner, L. A. Moustakas, and M. Giavalisco. Cosmic Variance in the Great Observatories Origins Deep Survey. *ApJL*, 600:L171–L174, January 2004. doi: 10.1086/378628.

R. S. Somerville, P. F. Hopkins, T. J. Cox, B. E. Robertson, and L. Hernquist. A semi-analytic model for the co-evolution of galaxies, black holes and active galactic nuclei. *MNRAS*, 391:481–506, December 2008. doi: 10.1111/j.1365-2966.2008.13805.x.

K. Spekkens, R. Giovanelli, and M. P. Haynes. The Cusp/Core Problem in Galactic Halos: Long-Slit Spectra for a Large Dwarf Galaxy Sample. *AJ*, 129: 2119–2137, May 2005. doi: 10.1086/429592.

D. N. Spergel, R. Bean, O. Doré, M. R. Nolta, C. L. Bennett, J. Dunkley, G. Hinshaw, N. Jarosik, E. Komatsu, L. Page, H. V. Peiris, L. Verde, M. Halpern, R. S. Hill, A. Kogut, M. Limon, S. S. Meyer, N. Odegard, G. S. Tucker, J. L. Weiland, E. Wollack, and E. L. Wright. Three-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Implications for Cosmology. *ApJS*, 170:377–408, June 2007. doi: 10.1086/513700.

V. Springel, S. D. M. White, A. Jenkins, C. S. Frenk, N. Yoshida, L. Gao, J. Navarro, R. Thacker, D. Croton, J. Helly, J. A. Peacock, S. Cole, P. Thomas, H. Couchman, A. Evrard, J. Colberg, and F. Pearce. Simulations of the formation, evolution and clustering of galaxies and quasars. *Nature*, 435:629–636, June 2005. doi: 10.1038/nature03597.

C. M. Springob, K. L. Masters, M. P. Haynes, R. Giovanelli, and C. Marinoni. SFI++. II. A New I-Band Tully-Fisher Catalog, Derivation of Peculiar Velocities, and Data Set Properties. *ApJS*, 172:599–614, October 2007. doi: 10.1086/519527.

D. V. Stark, S. S. McGaugh, and R. A. Swaters. A First Attempt to Calibrate the Baryonic Tully-Fisher Relation with Gas-Dominated Galaxies. *AJ*, 138:392–401, August 2009. doi: 10.1088/0004-6256/138/2/392.

S. Stierwalt, M. P. Haynes, R. Giovanelli, B. R. Kent, A. M. Martin, A. Saintonge, I. D. Karachentsev, and V. E. Karachentseva. The Arecibo Legacy Fast Alfa Survey. IX. The Leo Region H I Catalog, Group Membership, and the H I Mass Function for the Leo I Group. *AJ*, 138:338–361, August 2009. doi: 10.1088/0004-6256/138/2/338.

L. E. Strigari, J. S. Bullock, M. Kaplinghat, J. Diemand, M. Kuhlen, and P. Madau. Redefining the Missing Satellites Problem. *ApJ*, 669:676–683, November 2007. doi: 10.1086/521914.

R. A. Swaters and M. Balcells. The Westerbork HI survey of spiral and irregular galaxies. II. R-band surface photometry of late-type dwarf galaxies. *A&A*, 390: 863–878, August 2002. doi: 10.1051/0004-6361:20020449.

R. A. Swaters, R. Sancisi, T. S. van Albada, and J. M. van der Hulst. The rotation curves shapes of late-type dwarf galaxies. *A&A*, 493:871–892, January 2009. doi: 10.1051/0004-6361:200810516.

E. N. Taylor, A. M. Hopkins, I. K. Baldry, M. J. I. Brown, S. P. Driver, L. S. Kelvin, D. T. Hill, A. S. G. Robotham, J. Bland-Hawthorn, D. H. Jones, R. G. Sharp, D. Thomas, J. Liske, J. Loveday, P. Norberg, J. A. Peacock, S. P. Bamford, S. Brough, M. Colless, E. Cameron, C. J. Conselice, S. M. Croom, C. S. Frenk, M. Gunawardhana, K. Kuijken, R. C. Nichol, H. R. Parkinson, S. Phillipps, K. A. Pimbblet, C. C. Popescu, M. Prescott, W. J. Sutherland, R. J. Tuffs, E. van Kampen, and D. Wijesinghe. Galaxy And Mass Assembly (GAMA): stellar mass estimates. *MNRAS*, 418:1587–1620, December 2011. doi: 10.1111/j.1365-2966.2011.19536.x.

A. V. Tikhonov and A. Klypin. The emptiness of voids: yet another overabundance problem for the Λ cold dark matter model. *MNRAS*, 395:1915–1924, June 2009. doi: 10.1111/j.1365-2966.2009.14686.x.

J. L. Tinker and C. Conroy. The Void Phenomenon Explained. *ApJ*, 691:633–639, January 2009. doi: 10.1088/0004-637X/691/1/633.

C. Trachternach, W. J. G. de Blok, S. S. McGaugh, J. M. van der Hulst, and R.-J. Dettmar. The baryonic Tully-Fisher relation and its implication for dark matter halos. *A&A*, 505:577–587, October 2009. doi: 10.1051/0004-6361/200811136.

S. Trujillo-Gomez, A. Klypin, J. Primack, and A. J. Romanowsky. LCDM Correctly Predicts Basic Statistics of Galaxies: Luminosity-Velocity Relation,

Baryonic Mass-Velocity Relation, and Velocity Function. *ArXiv e-prints*, May 2010.

S. Trujillo-Gomez, A. Klypin, J. Primack, and A. J. Romanowsky. Galaxies in ΛCDM with Halo Abundance Matching: Luminosity-Velocity Relation, Baryonic Mass-Velocity Relation, Velocity Function, and Clustering. *ApJ*, 742:16, November 2011. doi: 10.1088/0004-637X/742/1/16.

D. Tweed, J. Devriendt, J. Blaizot, S. Colombi, and A. Slyz. Building merger trees from cosmological N-body simulations. Towards improving galaxy formation models using subhaloes. *A&A*, 506:647–660, November 2009. doi: 10.1051/0004-6361/200911787.

A. Vale and J. P. Ostriker. Linking halo mass to galaxy luminosity. *MNRAS*, 353: 189–200, September 2004. doi: 10.1111/j.1365-2966.2004.08059.x.

M. A. W. Verheijen and R. Sancisi. The Ursa Major cluster of galaxies. IV. HI synthesis observations. *A&A*, 370:765–867, May 2001. doi: 10.1051/0004-6361:20010090.

M. S. Warren, K. Abazajian, D. E. Holz, and L. Teodoro. Precision Determination of the Mass Function of Dark Matter Halos. *ApJ*, 646:881–885, August 2006. doi: 10.1086/504962.

S. M. Weinmann, F. C. van den Bosch, X. Yang, H. J. Mo, D. J. Croton, and B. Moore. Properties of galaxy groups in the Sloan Digital Sky Survey - II. Active galactic nucleus feedback and star formation truncation. *MNRAS*, 372: 1161–1174, November 2006. doi: 10.1111/j.1365-2966.2006.10932.x.

A. A. West, D. A. Garcia-Appadoo, J. J. Dalcanton, M. J. Disney, C. M. Rockosi, and Ž. Ivezić. H I-Selected Galaxies in the Sloan Digital Sky Survey. II. The Colors of Gas-Rich Galaxies. *AJ*, 138:796–807, September 2009. doi: 10.1088/0004-6256/138/3/796.

J. Wolf, G. D. Martinez, J. S. Bullock, M. Kaplinghat, M. Geha, R. R. Muñoz, J. D. Simon, and F. F. Avedo. Accurate masses for dispersion-supported galaxies. *MNRAS*, 406:1220–1237, August 2010. doi: 10.1111/j.1365-2966.2010.16753.x.

O. I. Wong, E. V. Ryan-Weber, D. A. Garcia-Appadoo, R. L. Webster, L. Staveley-Smith, M. A. Zwaan, M. J. Meyer, D. G. Barnes, V. A. Kilborn, R. Bhathal, W. J. G. de Blok, M. J. Disney, M. T. Doyle, M. J. Drinkwater, R. D. Ekers, K. C. Freeman, B. K. Gibson, S. Gurovich, J. Harnett, P. A. Henning, H. Jerjen, M. J. Kesteven, P. M. Knezek, B. S. Koribalski, S. Mader, M. Marquarding, R. F. Minchin, J. O'Brien, M. E. Putman, S. D. Ryder, E. M. Sadler, J. Stevens, I. M. Stewart, F. Stootman, and M. Waugh. The Northern HIPASS catalogue - data presentation, completeness and reliability measures. *MNRAS*, 371:1855–1864, October 2006. doi: 10.1111/j.1365-2966.2006.10846.x.

X. X. Xue, H. W. Rix, G. Zhao, P. Re Fiorentin, T. Naab, M. Steinmetz, F. C. van den Bosch, T. C. Beers, Y. S. Lee, E. F. Bell, C. Rockosi, B. Yanny, H. Newberg, R. Wilhelm, X. Kang, M. C. Smith, and D. P. Schneider. The Milky Way's Circular Velocity Curve to 60 kpc and an Estimate of the Dark Matter Halo Mass from the Kinematics of ˜2400 SDSS Blue Horizontal-Branch Stars. *ApJ*, 684:1143–1158, September 2008. doi: 10.1086/589500.

X. Yang, H. J. Mo, and F. C. van den Bosch. Galaxy Groups in the SDSS DR4. II. Halo Occupation Statistics. *ApJ*, 676:248–261, March 2008. doi: 10.1086/528954.

X. Yang, H. J. Mo, and F. C. van den Bosch. Galaxy Groups in the SDSS DR4. III. The Luminosity and Stellar Mass Functions. *ApJ*, 695:900–916, April 2009. doi: 10.1088/0004-637X/695/2/900.

N. Yasuda, M. Fukugita, and S. Okamura. Study of the Virgo Cluster Using the B-Band Tully-Fisher Relation. *ApJS*, 108:417–+, February 1997. doi: 10.1086/312960.

D. G. York, J. Adelman, J. E. Anderson, Jr., S. F. Anderson, J. Annis, N. A. Bahcall, J. A. Bakken, R. Barkhouser, S. Bastian, E. Berman, W. N. Boroski, S. Bracker, C. Briegel, J. W. Briggs, J. Brinkmann, R. Brunner, S. Burles, L. Carey, M. A. Carr, F. J. Castander, B. Chen, P. L. Colestock, A. J. Connolly, J. H. Crocker, I. Csabai, P. C. Czarapata, J. E. Davis, M. Doi, T. Dombeck, D. Eisenstein, N. Ellman, B. R. Elms, M. L. Evans, X. Fan, G. R. Federwitz, L. Fiscelli, S. Friedman, J. A. Frieman, M. Fukugita, B. Gillespie, J. E. Gunn, V. K. Gurbani, E. de Haas, M. Haldeman, F. H. Harris, J. Hayes, T. M. Heckman, G. S. Hennessy, R. B. Hindsley, S. Holm, D. J. Holmgren, C.-h. Huang, C. Hull, D. Husby, S.-I. Ichikawa, T. Ichikawa, Ž. Ivezić, S. Kent, R. S. J. Kim, E. Kinney, M. Klaene, A. N. Kleinman, S. Kleinman, G. R. Knapp, J. Korienek, R. G. Kron, P. Z. Kunszt, D. Q. Lamb, B. Lee, R. F. Leger, S. Limmongkol, C. Lindenmeyer, D. C. Long, C. Loomis, J. Loveday, R. Lucinio, R. H. Lupton, B. MacKinnon, E. J. Mannery, P. M. Mantsch, B. Margon, P. McGehee, T. A. McKay, A. Meiksin, A. Merelli, D. G. Monet, J. A. Munn, V. K. Narayanan, T. Nash, E. Neilsen, R. Neswold, H. J. Newberg, R. C. Nichol, T. Nicinski, M. Nonino, N. Okada, S. Okamura, J. P. Ostriker, R. Owen, A. G. Pauls, J. Peoples, R. L. Peterson, D. Petravick, J. R. Pier, A. Pope, R. Pordes, A. Prosapio, R. Rechenmacher, T. R. Quinn, G. T. Richards, M. W. Richmond, C. H. Rivetta, C. M. Rockosi, K. Ruthmansdorfer, D. Sandford, D. J. Schlegel, D. P. Schneider, M. Sekiguchi, G. Sergey, K. Shimasaku, W. A. Siegmund, S. Smee, J. A. Smith, S. Snedden, R. Stone, C. Stoughton, M. A. Strauss, C. Stubbs, M. SubbaRao, A. S. Szalay, I. Szapudi, G. P. Szokoly, A. R. Thakar, C. Tremonti, D. L. Tucker, A. Uomoto, D. Vanden Berk, M. S. Vogeley, P. Waddell, S.-i. Wang, M. Watanabe, D. H. Weinberg, B. Yanny, N. Yasuda, and SDSS Collaboration. The Sloan Digital

Sky Survey: Technical Summary. *AJ*, 120:1579–1587, September 2000. doi: 10.1086/301513.

H. J. Zahid, G. I. Dima, L. J. Kewley, D. K. Erb, and R. Davé. A Census of Oxygen in Star-forming Galaxies: An Empirical Model Linking Metallicities, Star Formation Rates, and Outflows. *ApJ*, 757:54, September 2012. doi: 10.1088/0004-637X/757/1/54.

J. Zavala, Y. P. Jing, A. Faltenbacher, G. Yepes, Y. Hoffman, S. Gottlöber, and B. Catinella. The Velocity Function in the Local Environment from ΛCDM and ΛWDM Constrained Simulations. *ApJ*, 700:1779–1793, August 2009. doi: 10.1088/0004-637X/700/2/1779.

I. Zehavi, Z. Zheng, D. H. Weinberg, J. A. Frieman, A. A. Berlind, M. R. Blanton, R. Scoccimarro, R. K. Sheth, M. A. Strauss, I. Kayo, Y. Suto, M. Fukugita, O. Nakamura, N. A. Bahcall, J. Brinkmann, J. E. Gunn, G. S. Hennessy, Ž. Ivezić, G. R. Knapp, J. Loveday, A. Meiksin, D. J. Schlegel, D. P. Schneider, I. Szapudi, M. Tegmark, M. S. Vogeley, D. G. York, and SDSS Collaboration. The Luminosity and Color Dependence of the Galaxy Correlation Function. *ApJ*, 630:1–27, September 2005. doi: 10.1086/431891.

I. Zehavi, Z. Zheng, D. H. Weinberg, M. R. Blanton, N. A. Bahcall, A. A. Berlind, J. Brinkmann, J. A. Frieman, J. E. Gunn, R. H. Lupton, R. C. Nichol, W. J. Percival, D. P. Schneider, R. A. Skibba, M. A. Strauss, M. Tegmark, and D. G. York. Galaxy Clustering in the Completed SDSS Redshift Survey: The Dependence on Color and Luminosity. *ApJ*, 736:59, July 2011. doi: 10.1088/0004-637X/736/1/59.

W. Zhang, C. Li, G. Kauffmann, H. Zou, B. Catinella, S. Shen, Q. Guo, and R. Chang. Estimating the HI gas fractions of galaxies in the local Universe. *MNRAS*, 397:1243–1253, August 2009. doi: 10.1111/j.1365-2966.2009.15050.x.

M. A. Zwaan, L. Staveley-Smith, B. S. Koribalski, P. A. Henning, V. A. Kilborn, S. D. Ryder, D. G. Barnes, R. Bhathal, P. J. Boyce, W. J. G. de Blok, M. J. Disney, M. J. Drinkwater, R. D. Ekers, K. C. Freeman, B. K. Gibson, A. J. Green, R. F. Haynes, H. Jerjen, S. Juraszek, M. J. Kesteven, P. M. Knezek, R. C. Kraan-Korteweg, S. Mader, M. Marquarding, M. Meyer, R. F. Minchin, J. R. Mould, J. O'Brien, T. Oosterloo, R. M. Price, M. E. Putman, E. Ryan-Weber, E. M. Sadler, A. Schröder, I. M. Stewart, F. Stootman, B. Warren, M. Waugh, R. L. Webster, and A. E. Wright. The 1000 Brightest HIPASS Galaxies: The H I Mass Function and Ω_{H_I}. *AJ*, 125:2842–2858, June 2003. doi: 10.1086/374944.

M. A. Zwaan, M. J. Meyer, R. L. Webster, L. Staveley-Smith, M. J. Drinkwater, D. G. Barnes, R. Bhathal, W. J. G. de Blok, M. J. Disney, R. D. Ekers, K. C. Freeman, D. A. Garcia, B. K. Gibson, J. Harnett, P. A. Henning, M. Howlett, H. Jerjen, M. J. Kesteven, V. A. Kilborn, P. M. Knezek, B. S. Koribalski, S. Mader,

M. Marquarding, R. F. Minchin, J. O'Brien, T. Oosterloo, M. J. Pierce, R. M. Price, M. E. Putman, E. Ryan-Weber, S. D. Ryder, E. M. Sadler, J. Stevens, I. M. Stewart, F. Stootman, M. Waugh, and A. E. Wright. The HIPASS catalogue - II. Completeness, reliability and parameter accuracy. *MNRAS*, 350:1210–1219, June 2004. doi: 10.1111/j.1365-2966.2004.07782.x.

M. A. Zwaan, M. J. Meyer, L. Staveley-Smith, and R. L. Webster. The HIPASS catalogue: Ω_{H_I} and environmental effects on the HI mass function of galaxies. *MNRAS*, 359:L30–L34, May 2005. doi: 10.1111/j.1745-3933.2005.00029.x.

M. A. Zwaan, M. J. Meyer, and L. Staveley-Smith. The velocity function of gas-rich galaxies. *MNRAS*, 403:1969–1977, April 2010. doi: 10.1111/j.1365-2966.2009.16188.x.